

# DI NEREIKIA JAUSTI. JAM PAKANKA IŠMOKTI, NUO KO JAUČIAME MES.

Žmogus į DI pasaulį ateina ne „tuščias“. Kultūra jau yra suformavusi daugybę asociacijų: vėliava gali sukelti priklausymo jausmą, himnas – pakilimą, priešų vaizdas – grėsmę, tam tikras balsas – autoriteto įspūdį, minia – vienybės jausmą, muzikinė kulminacija – didybę.

**Ne visiems. Ne visada. Ne vienodai. Bet mūsų reakcijos yra pakankamai dėsningos, kad jų dėsningumus būtų galima mokytis prognozuoti.**

**Ir štai čia atsiranda DI. Jam nereikia mylėti Tėvynės, bijoti priešų, patirti didybės ar apskritai ką nors jausti. Jam pakanka išmokti, kokia žodžių, vaizdų, muzikos, balso ir simbolių kombinacija padidina tikimybę, kad JAUSI TU.**

Ir tai jau ne fantastika.

2025 m. Nature Communications paskelbtame tyrime trijuose iš anksto registruotuose eksperimentuose dalyvavo 4829 žmonės. LLM sugeneruotos politinės žinutės reikšmingai keitė žmonių požiūrį į įvairius politikos klausimus. Ir čia svarbus net ne pats poveikio dydis. Žmogaus nereikia „užvaldyti“. Pakanka truputį pakeisti tikimybę, kad jis pasirinks vieną interpretaciją, o ne kitą. O tada tą mažą pranašumą galima dauginti mastu – milijonams žmonių, milijonus kartų.

2024 m. Scientific Reports paskelbtame keturių tyrimų cikle su 1788 dalyviais ChatGPT sukurtos, pagal žmogaus psichologines savybes pritaikytos žinutės buvo įtaigesnės už nepritaikytas. O 2025 m. Nature paskelbti eksperimentai realiuose JAV, Kanados ir Lenkijos rinkimų kontekstuose parodė, kad pokalbiai su DI, argumentuojančiu kandidatų naudai, reikšmingai keitė dalyvių kandidatų vertinimus. Tai nėra „proto kontrolė“, bet tai jau eksperimentiškai išmatuojamas politinio įtikinėjimo poveikis.

## Ir dabar prasideda įdomiausia.

**DI pranašumas nėra tas, kad žmogus yra kvailas. DI pranašumas tas, kad didelė dalis mūsų reakcijų yra tipinės.**

Žmogus pamato situaciją ir dažnai nepradeda jos analizuoti nuo nulio. Jis atpažįsta kategoriją: „savas“, „svetimas“, „geras“, „blogas“, „mūsų“, „jų“, „auka“, „priešas“. O kartu su kategorija jau ateina ir išmoktas santykis bei beveik paruošta reakcija.

**Schema paprasta: situacija — atpažinta kategorija — išmoktas santykis — reakcija.**

Tai labai sena žmogaus savybė. Bet DI atsiradimas ją pakeičia iš esmės, nes mašinai lengviausia modeliuoti būtent tai, kas kartojasi. DI nereikia suprasti tavo „sielos“. Jam gali pakakti pastebėti, kad tam tikram žmogui, tam tikrame kontekste, tam tikras žodis, vaizdas, balsas ar simbolis statistiškai dažniau sukelia tam tikrą reakciją.

**Ir kuo reakcija tipiškesnė, tuo ji lengviau prognozuojama.**

**Štai čia „mechaninio žmogaus“ metafora pradeda skambėti visai kitaip. Mechaninis žmogus nėra tas, kuris naudojasi mašina. Mechanizacija prasideda tada, kai pats žmogaus reagavimas tampa toks šabloniškas, kad mašina gali vis tiksliau jį modeliuoti.**

Kuo daugiau žmogus reaguoja tiesiogiai į konkrečią tikrovę, tuo sunkiau jį sutalpinti į modelį. Kuo daugiau jis reaguoja pagal jau paruoštą „čia mūsų“, „čia jų“, „čia didinga“, „čia šventa“, „čia pavojinga“, „čia reikia pykti“, tuo jo reakcija tampa labiau nuspėjama.

Ir tada senoji propaganda atrodo beveik primityvi. Ji kūrė vieną spektaklį tūkstančiams žmonių: vieną plakatą, vieną kalbą, vieną himną, vieną priešą. DI potencialiai leidžia apversti formulę – tūkstančiui žmonių sukurti tūkstantį skirtingų to paties spektaklio versijų.

Vienam parodyti tai, kas sužadina baimę, kitam – pasididžiavimą, trečiam – nostalgiją, ketvirtam – pyktį, penktam – priklausymo „mūsiškiams“ jausmą.

**Ir nereikia žmogaus paversti marionete. Pakanka statistiškai padidinti tikimybę, kad tam tikru momentu jo reakcija pasislinks norima kryptimi.**

Dar įdomiau tai, kad žmogus ir DI gali pradėti mokytis vienas kitą. Nature Human Behaviour paskelbtame eksperimente su 1401 dalyviu nustatyta, kad sąveika su šališku DI gali sustiprinti žmonių perceptinius, emocinius ir socialinius vertinimo šališkumus, o patys žmonės ne visada pakankamai įvertina, kiek DI paveikė jų sprendimus.

Taigi ciklas tampa labai paprastas: kultūra suformuoja reakcijų kanalus — DI juos aptinka — pateikia stimulą — žmogus sureaguoja — sistema gauna grįžtamąjį ryšį — prisitaiko — kitas stimulus tampa tikslesnis.

Ir žmogus sako: „Bet juk AŠ taip jaučiu.“

Taip. Emocija tikrai tavo. Pyktis tavo. Baimė tavo. Pakilimas tavo. Pasididžiavimas tavo.

Bet tai dar neatsako į svarbiausią klausimą:

**KAS IR KAIP SUKŪRĖ SĄLYGAS, KAD BŪTENT ŠI EMOCIJA KILO BŪTENT DABAR?**

**Čia ir matau tikrąją problemą. Ne tai, kad DI gali tapti panašus į žmogų, o tai, kad DI gali vis geriau išmokti žmogų kaip reakcijų sistemą.**

Žmogus nėra šachmatų lenta. Jo neįmanoma apskaičiuoti iki paskutinio ėjimo. Kūno būseną, gyvenimo patirtis, aplinka, santykiai, nauja informacija ir sąmoningas apsisprendimas viską keičia. Bet DI ir nereikia mūsų apskaičiuoti iki galo. Jam pakanka būti šiek tiek geresniam už atsitiktinumą – ir tą pranašumą dauginti mastu, greičiu, kartojimu bei personalizacija.

Štai kur „mechaninio žmogaus“ metafora ima skambėti nemaloniai šiuolaikiškai: mašina darosi vis panašesnė į žmogų, o tipiškos žmogaus reakcijos – vis lengviau paverčiamos modeliu.

**DI neturi emocijų. Bet jam jų ir nereikia.**

**Jam pakanka išmokti mūsų emocijų klaviatūrą.**

**Ir pavojingiausia gali būti ne diena, kai mašina išmoks jausti kaip žmogus. Pavojingiausia bus diena, kai žmogus nebesupras, kiek jo „spontaniško“ jausmo buvo sužadinta iš išorės.**

Jis vis tiek visiškai nuoširdžiai sakys:

**„Bet juk AŠ taip jaučiu.“**

**Taip.**

**Tik kas tuo metu groja tavo klaviatūra?**