

DI KEIČIA SISTEMĄ. BET KĄ JOJE NORIME IŠSAUGOTI?

Diskusijose apie dirbtinį intelektą nuolat kartojama ta pati schema.

DI bus naudingas arba pavojingas.

Sumažins nelygybę arba ją padidins.

Padės žmogui arba jį pakeis.

Liks kontroliuojamas arba kontrolę prarasime.

Panašiai klausimą vienoje naujausių savo esė formuluoja ir Billas Gatesas.

Tačiau problema gerokai gilesnė.

DI PATS SAVAIME NĖRA NEI LYGYBĖ, NEI NELYGYBĖ.

Tas pats technologinis pajėgumas vienoje santykių sistemoje gali mažinti skirtumus, o kitoje – juos didinti.

Todėl klausimas „ar DI bus geras, ar blogas?“ yra daug silpnesnis už kitą:

KOKIUS SANTYKIUS DI PAKEIS?

Čia prasideda tikroji problema.

Dabartinė ekonomika nemaža dalimi remiasi ryšiais:

darbas – pajamos – vartojimas

ir

pajamos – mokesčiai – viešosios paslaugos bei socialinė apsauga.

Tai, žinoma, nėra visas ekonomikos modelis. Egzistuoja kapitalo pajamos, nuosavybė, kreditai, transferai ir daugybė kitų ryšių.

Tačiau milijardams žmonių darbas vis dar yra pagrindinis būdas gauti pajamas.

Kas atsitinka, jeigu DI ir robotizacija leidžia gaminti vis daugiau, o žmogaus darbo tam reikia vis mažiau?

Problema tada nėra vien „žmonės neteks darbo“.

Keistis pradeda pats santykis:

DARBAS – PAJAMOS.

O kartu kyla daug fundamentalesni klausimai.

Kaip žmogus gauna prieigą prie sistemos sukurtų gėrybių?

Kas finansuoja valstybę?

Iš kur atsiranda masinė perkamoji galia?

Kas nutinka ekonomikai, jeigu produktyvumas didėja, tačiau gamybai reikia vis mažiau tų pačių žmonių, kurie kartu yra jos vartotojai?

Tai jau ne vien darbo rinkos problema.

Tai santykių sistemos transformacija.

Ir čia atsiranda klausimas, kuris, mano manymu, yra svarbesnis už beveik visas diskusijas apie DI reguliavimą:

KĄ BŪTENT NORIME IŠSAUGOTI?

Žmogaus gyvybę, materialinį saugumą, orumą, autonomiją, atsakomybę? Galimybę pačiam nustatyti savo tikslus ir dalyvauti visuomenės sukurta rezultate?

O gal norime išsaugoti ryšį „darbas – pajamos“, rinkos ekonomiką, privačios nuosavybės struktūrą, dabartines institucijas ar tiesiog socialinį stabilumą?

Tai nėra tas pats.

Ir čia slypi esminė skirtis:

IŠSAUGOTI ŽMOGŲ NĖRA TAS PATS, KAS IŠSAUGOTI SISTEMĄ, KURIOJE JIS IKI ŠIOL GYVENO.

Sistema gali pasikeisti beveik neatpažįstamai ir vis tiek išsaugoti tai, kas žmogui esminga.

Ir priešingai – sistema gali išoriškai beveik nepasikeisti, bet žmogus joje gali prarasti būtent tai, ką norėjome apsaugoti.

Todėl stabilumas pats savaime nėra kriterijus.

Stabili gali būti ir labai bloga sistema.

Reikia klausti:

KAS TURI IŠLIKTI, NET JEIGU VISA SISTEMA PASIKEIS?

Matematikoje savybę, kuri išlieka vykstant transformacijai, vadiname invariantu.

Galbūt žmogaus atveju tai net nėra viena savybė. Gal turime ieškoti viso invariantų rinkinio – tokių dalykų, kuriuos laikome esminiais nepriklausomai nuo to, kokią formą įgaus ekonomika ar technologijos.

Todėl vienas svarbiausių DI eros klausimų būtų:

KOKS YRA ŽMOGAUS INVARIANTAS?

Pavyzdžiui, dabar daug kalbama apie profesijas, kurias galbūt vertėtų palikti žmogui net tada, kai jas techniškai galėtų atlikti DI ar robotas.

Bet čia reikia atskirti du visiškai skirtingus dalykus.

Galbūt tam tikras žmogus – žmogus santykis pats savaime turi vertę – globoje, ugdyme, slaugoje, bendraujant su vaiku ar lydint mirštantį žmogų.

Tada išsaugome ne „darbo vietą“, o tam tikro tipo žmogišką santykį.

Tačiau galima žmogaus darbą saugoti ir dėl visiškai kitos priežasties – todėl, kad dabartinė ekonominė sistema vis dar didžiąja dalimi reikalauja ryšio:

darbas – pajamos.

Tokiu atveju galime saugoti ne žmogų, o tiesiog dirbtinai palaikyti seną ekonominį mechanizmą.

Todėl klausimas labai paprastas:

AR SAUGOME ŽMOGŲ – AR DARBO BŪTINYBĘ?

Jeigu žmogaus darbas buvo būtinas todėl, kad be jo nebuvo galima pagaminti rezultato, o technologija šią būtinybę iš esmės susilpnina ar kada nors panaikina, nebūtinai turime amžinai saugoti seną santykį.

Gal reikia sukurti kitą.

Dar viena nuolat kartojama frazė:

„DI pranoks žmogų.“

Bet pranoks pagal ką?

Skaičiavimo greitį, atmintį, programavimą, diagnostiką, matematiką, strateginį planavimą, hipotezių generavimą?

O kaip dėl vertės pasirinkimo, patyrimo ar sąmonės?

Be kriterijaus sakiny *„DI pranoks žmogų“* suspaudžia daugybę skirtingų funkcijų į vieną įsivaizduojamą skalę.

Kalkuliatorius žmogų jau seniai pranoksta aritmetikos greičiu.

Tačiau iš to neišplaukia:

kalkuliatorius > žmogus.

Lokalių funkcijų pranašumas nėra viso objekto pranašumo įrodymas.

Todėl daug adekvatesnis klausimas yra ne vien apie gebėjimus.

O apie sisteminę autonomiją.

Čia sistemine autonomija vadinu ne kažkokią absoliučiai nuo pasaulio nepriklausomą sistemą – tokios nėra. Kalbu apie daug konkretesnį kriterijų: ar technologinė sistema gali pati palaikyti vis didesnę dalį būtinų savo veikimo ir reprodukcijos sąlygų be būtinos žmogaus operacijos?

Šiandien net pats galingiausias DI priklauso nuo milžiniškos žmogaus sukurtos infrastruktūros – elektros tinklų, duomenų centrų, lustų fabrikų, žaliavų gavybos, logistikos, robotikos, remonto, programinės infrastruktūros, institucijų ir žmonių, kurie visa tai palaiko.

Todėl faktas, kad DI pats parašė programą, atliko tyrimą ar suplanavo sudėtingą užduotį, dar nereiškia, kad atsirado autonomiška sistema.

GEBĖJIMAS NĖRA AUTONOMIJA.

Ir kalbėdamas apie *„DI sistemą“* čia turiu omenyje jau ne vien kalbos modelį kompiuteryje, o platesnį technologinį kompleksą – DI, robotiką, energetiką, gamybą, logistiką ir infrastruktūrą.

Tikras struktūrinis lūžis būtų visai kitur.

Tarkime, tokia technologinė sistema galėtų vis labiau pati palaikyti būtiną savo reprodukcijos kontūrą:

ištekliai - energija - gamyba - infrastruktūros priežiūra - remontas - atnaujinimas - nauji ištekliai.

Ir žmogaus operacija tam tikrame tokiam uždaramame kontūre nebebūtų būtina.

Tada jau turėtume ne tiesiog *„labai protingą DI“*.

Turėtume sistemą, kuri pati atkuria tam tikras būtinas savo veikimo sąlygas.

Ir tuomet *„kontrolės praradimas“* įgytų daug tikslesnę reikšmę.

Ne:

„mašina tapo protingesnė“.

O:

ŽMOGUS NUSTOJO BŪTI BŪTINA TAM TIKRO SISTEMOS REPRODUKCIJOS KONTŪRO GRANDIS.

Šį teiginį jau galima tikrinti.

Kur žmogus vis dar būtinas? Kur sistema be jo neužsidaro? Kokias savo veikimo sąlygas ji jau gali atkurti pati? Ar mažėja funkcijų, kurioms žmogus vis dar būtinas?

Tai daug rimtesni klausimai už abstraktų *„ar DI taps protingesnis už mus?“*

Ir tada grįžtame į pradžią.

DI gali radikaliai pakeisti darbą, pajamas, švietimą, mediciną, institucijas, nuosavybės santykius, žmogaus vaidmenį gamyboje ir net žmogaus santykį su pačiu sprendimų priėmimu.

Tačiau jeigu visa tai gali keistis, mums reikia žinoti ne tik:

KAS KEISIS?

Daug svarbiau:

KAS NETURI PASIKEISTI?

Jeigu invariantas yra žmogaus autonomija – tikrinkime autonomiją.

Jeigu orumas – apibrėžkime, kas jį pažeidžia.

Jeigu materialinis saugumas – tuomet nebūtina amžinai išsaugoti konkretaus mechanizmo „darbas – pajamos“, jeigu tą patį invariantą galima užtikrinti kitu santykiu.

Jeigu žmogaus galimybė pačiam nustatyti tikslus – tikrinkime, ar DI lieka priemone tiems tikslams, ar vis labiau pradeda dalyvauti nustatant ir pačius tikslus.

Štai tada galima prasmingai kalbėti apie DI saugumą.

Ne vien apie senos sistemos stabilumą, o apie pasirinkto invarianto išsaugojimą sistemai transformuojantis.

Ir čia atsiranda paradoksas.

Sistema gali pasikeisti beveik neatpažįstamai ir išsaugoti tai, kas žmogui svarbiausia.

O gali likti beveik tokia pati – ir prarasti būtent tai, dėl ko ją apskritai buvo verta saugoti.

Todėl pagrindinis DI eros klausimas man nėra:

AR DI BUS GERAS, AR BLOGAS?

Net ne:

AR DI PRANOKS ŽMOGŲ?

Ir net ne:

AR MES JĮ KONTROLIUOSIME?

Pirmasis klausimas turėtų būti:

KĄ ŽMOGUJE IR ŽMOGAUS SANTYKIUOSE MES LAIKOME TOKIA VERTE, KURI TURI IŠLIKTI NET TADA, JEIGU VISA APLINK JĄ ESANTI SISTEMA PASIKEIS?

Kol į tai neatsakėme, mes dar net nežinome, ką bandome apsaugoti nuo DI.

Gal žmogų.

O gal tik vakarykštę sistemą, kurios išlikimą jau painiojame su paties žmogaus išlikimu.