

„Asinchroninis dirbtinių intelektų konsiliumo metodas (ADIKM)“

arba

Asinchroninis DI konsiliumas kaip sprendimų priėmimo architektūra neapibrėžtumo sąlygomis

Saulius Lapienis

Vilniaus universitetas

Duomenų mokslo ir skaitmeninių technologijų institutas

2026 vasario 23 d.

Mano technologinė trajektorija

1972–1983

Minsk-22A, BESM-6, FORTRAN

Pirmieji programavimo metai – didžiosios mašinos ir 0/1 logika.

1987–1995

IBM PC XT, DOS, Basic, Framework IV, SuperCalc, Excel 2.0, pirmieji serveriai, el. pašto pradžia Lietuvoje.

1991–2010

386, 486, Pentium ir t.t, Windows evoliucija

Asmeninių kompiuterių ir tinklų augimas.

2005–2006

Lygiagretūs skaičiavimai, superkompiuteris.

2019–...

Kvantinių skaičiavimų bandymai.

Dirbtinis intelektas kaip nauja architektūrinė fazė.

Perėjau kelias technologijų bangas.

DI man kaip dar viena sisteminė transformacija.

Kodėl ši tema reikalinga

- Sprendimus vis dažniau priima ne vienas DI
- Skirtingi DI turi skirtingus duomenis ir laiką
- Pilna sinchronizacija realybėje neįmanoma

- Pagrindinis klausimas:

Kaip priimti sprendimą be dirbtinės sinchronizacijos?

Kas yra DI konsiliumas

DI konsiliumas – tai kelių nepriklausomų DI sistemų dalyvavimas sprendime

Tai ne:

- "*ensemble learning*"
- balsavimas
- vidurkis

Tai:

sprendimų architektūra, o ne algoritmas

Kodėl sinchronizacija yra iliuzija

- Duomenys ateina skirtingu metu
- Modeliai mokytis skirtingais laikotarpiais
- Kontekstai niekada pilnai nesutampa

Priverstinė sinchronizacija slepia konfliktus

Asinchroniškumas kaip privalumas

- Išsaugo nesutarimus
- Atveria neapibrėžtumą
- Atskiria tvirtus teiginius nuo abejonių

Sprendimas gimsta iš struktūruoto konflikto

Ryšys su paskirstytomis sistemomis

Sinchroninės sistemos:

- Bendras laikrodis
- Greitas konsensusas
- Trapumas

Asinchroninės sistemos:

- Nėra bendro laiko
- Dalinis konsensusas
- Atsparumas

Kodėl vidurkinimas neveikia

- Sunaikina kraštutinius, bet svarbius „signalus“
- Supainioja nežinojimą su nesutikimu
- Pašalina laiko dimensiją

Siūloma konsiliumo struktūra

1. Nepriklausomi DI atsakymai
2. Laiko žymos
3. Pozicijų tipai (teiginys, hipotezė, abejonė)
4. Konfliktų žemėlapis
5. Žmogaus arba meta-DI sprendimas

Metodo idėja:

DI konsiliumas kaip architektūra

DI konsiliumas – tai ne algoritmas,
o sprendimų priėmimo architektūra.

Pagrindinė idėja:

- keli nepriklausomi DI veikia lygiagrečiai
- jų atsakymai nėra sinchronizuojami prievarta
- svarbūs ne tik atsakymai, bet ir jų skirtumai

Architektūros komponentai

1. Užklausa (problemos formulavimas)
2. Nepriklausomi DI agentai
3. Atsakymų registras (su laiko žymomis)
4. Konfliktų / nesutarimų analizė
5. Sprendimo priėmėjas (žmogus arba meta-DI)

Asinchroninio veikimo principas

- DI atsako skirtingu metu
- atsakymai gali prieštarauti
- nėra reikalavimo pasiekti konsensą

Svarbu:

nesutarimas laikomas informacija, o ne klaida

Minimalus pavyzdys (1)

Užklausa:

„Ar situacija X kelia didelę riziką?“

DI A: „Taip, aukšta rizika“ (remiasi istorine statistika)

DI B: „Ne, rizika žema“ (remiasi dabartiniais duomenimis)

DI C: „Nežinau, trūksta duomenų“

Minimalus pavyzdys (2) – sprendimo logika

Konsiliumas NEIEŠKO vidurkio.

Jis pateikia struktūrą:

- kur yra konfliktas
- kas remiasi praeitimi, kas dabartimi
- kas pripažįsta nežinojimą

Sprendimą priima žmogus arba meta-DI, matydamas visą paveikslą

Santykis su techniniais DI tyrimais

- Šis darbas veikia virš modelių
- Nenurodo, kaip modelis mokomas
- Nurodo, kaip skirtingi modeliai jungiami

Tai papildantis, o ne konkuruojantis lygmuo

Kur tai jau aktualu šiandien

- Kibernetinis saugumas
- Geopolitinė analizė
- Medicina
- Krizių valdymas
- DI sauga

Darbo ribos

Šis pranešimas:

- ✘ Nesiūlo naujo algoritmo
- ✘ Neoptimizuoja „accuracy“ ar „loss“
- ✘ Nekonkuruoja su esamais modeliais
- ✓ Siūlo sprendimų architektūrą

Techninis gylis ir atsakomybė

Techniniai sprendimai:

- Paliekami konkrečių metodų autoriams
- Gali būti skirtingi skirtingose srityse

Architektūra:

- Leidžia jiems veikti kartu

Atviri klausimai – kvietimas bendradarbiauti

- Formalus aprašymas
- Metrikos
- Empiriniai tyrimai

Tai sąmoningai paliekama galimam bendram darbui

Empirinis kontekstas

Ši architektūra kilo ne vien teoriškai.

- Buvo atliktas 16 nepriklausomų DI sistemų konsiliumas
- Atsakymai buvo asinchroniniai ir prieštaringi
- Nebuvo priverstinio konsensuso

Šis pranešimas – konceptualus praktinės patirties apibendrinimas.

Išvada

Asinchroninis DI konsiliumas:

- Nėra silpnesnis DI
- Yra aukštesnio lygmens architektūra
- Būtinai realiame, neapibrėžtame pasaulyje

Nuo konsiliumo prie audito

- Konsiliumas mažina klaidos tikimybę *sprendimo momentu*
- Auditas tikrina klaidų kilmę *po sprendimo*
- Abi struktūros reikalingos

Žinios apie DI auditą jau vystosi, bet situacija yra *kūrybos stadijoje* ir dar nėra „baigto pasaulinio mechanizmo“.

Vienas žymiausių politikos ekspertų **Miles Brundage** įkūrė organizaciją, skirtą **nepriklausomam „frontier AI“ modelių auditui** — kitaip sakant, kad DI kūrėjai *ne patys vertintų savo saugumą*, o tai darytų išorės auditoriai.



NeurIPS - Conference on Neural Information Processing Systems



<https://www.youtube.com/@horizonsdelia>

„NeurIPS 2026“ - 2026 gruodžio 6-12, Sidnėjus, Australija

125 NEXT-GEN AI Innovations



Baigiant mintys aptarimui...

DI konsiliumas prasideda ne nuo modelių pasirinkimo,
o nuo klausimo:

kurie DI apskritai turi epistemines teises dalyvauti
sprendime

Post Scriptum:

Vienos konkrečios realizacijos pavyzdys aprašytas straipsnelyje:

<https://www.aviacijospasaulis.lt/straipsniai/naujienos-02/2025/11/sesiolika-dirbtiniu-intelektu-viena-issvada-lietuvai-skaidrumas-atsakomybe-ir-maziausia-butina-jega>

Asinchroninis DI konsiliumas kaip sprendimų priėmimo architektūra neapibrėžtumo sąlygomis

Parengė DI „Litua“ (ChatGPT 5.2) Sauliaus Lapienio
2026 02 23 d. pranešimo „**Asinchroninis dirbtinių intelektų konsiliumo metodas (ADIKM)**“ pagrindu.

Problemos formulė

Šiuolaikinė DI sistema:

- generuoja atsakymą;
- atrodo užtikrinta;
- neturi episteminių stabdžių.

Klausimas:

Kaip sumažinti klaidos tikimybę be centrinės autoriteto struktūros?

Vieno modelio riba

Vienas modelis = viena trajektorija.

Net jei:

- tikslumas 95 %;
- klaidos tikimybė 5 %

Sudėtingoje sistemoje:

5 % tampa sisteminė rizika.

Konsiliumo principas

Medicina:

- vienas gydytojas → nuomonė;
- konsiliumas → sprendimo kokybės augimas.

Perkėlimas į DI:

Ne vienas modelis, o asinchroninė daugiamodelinė struktūra.

Asinchroniškumas

Sinchroninis balsavimas \neq konsiliumas.

Asinchroninis modelis:

- modeliai nežino vienas kito atsakymų;
- atsakymai generuojami nepriklausomai;
- palyginimas vyksta po fakto

Tai sumažina:

- kolektyvinę klaidą;
- „echo efektą“

Episteminis „Stop“ principas

Pirmas klausimas sistemoje turi būti ne „atsakyk“, o „Patikrink statusą.“

Jei:

- modelių išvados diverguoja;
- pasitikėjimo lygis žemas

Sistema įjungia STOP režimą.

Matematinis aspektas

Jei modeliai nepriklausomi:

$P(\text{klaida konsiliume}) < P(\text{klaida viename modelyje})$

Net ir dalinė nepriklausomybė mažina bendrą riziką.

Tai artima:

- Bayes agregacijai
- klaidos slopinimui per redundanciją
- triukšmo filtravimui

Geopolitinis kontekstas

„Frontier“ modeliai tampa:

- infrastruktūra;
- sprendimų priėmimo dalimi;
- informacinės galios įrankiu

Todėl reikalingas:

struktūrinis auditavimo mechanizmas.

AI Auditas vs AI Konsiliumas

AI Auditas:

- „ex post“ patikrinimas;
- atitiktis reglamentams

AI Konsiliumas:

- „ex ante“ klaidų slopinimas;
- struktūrinė klaidos mažinimo architektūra

Rizikos mažinimo architektūra

Vienas modelis → deterministinė trajektorija

Konsiliumas → tikimybinė sprendimo stabilizacija

Ribos

- Modeliai gali turėti koreliuotas klaidas;
- Reikia nepriklausomų architektūrų;
- Reikia skirtingų mokymo šaltinių.

Išvada

Asinchroninis DI konsiliumas nėra:

- nuomonės demokratija;
- balsavimas

Tai yra:

struktūrinė klaidos tikimybės mažinimo schema.