

**NEPRIKLAUSOMA  
AUKŠTO LYGIO  
EKSPERTŲ GRUPĖ DIRBTINIO INTELEKTO  
KLAUSIMAI**

**EUROPOS KOMISIJOS SUDARYTA 2018 M. BIRŽELIO MĖN.**



**DI APIBRĖŽTIS.  
PAGRINDINIAI PAJĖGUMAI IR MOKSLO  
ŠAKOS**

**Apibrėžtis skirta vartoti Grupės rengiamuose dokumentuose**

# DI apibrėžtis.

## Pagrindiniai pajėgumai ir mokslo šakos

### Aukšto lygio ekspertų grupė DI klausimais

**Atsakomybės apribojimas ir šio dokumento naudojimas.** Toliau pateiktame DI pajėgumų ir mokslo šakų apibūdinime ir apibrėžtyje pažangiausi metodai labai supaprastinti. Šiame dokumente yra siekiama ne tiksliai ir išsamiai apibrėžti DI metodus ir pajėgumus, bet glaustai aprašyti, kokiais šios mokslo šakos aiškinimo principais vadovaujasi Aukšto lygio ekspertų grupė rengdama savo dokumentus. Tačiau tikimės, kad žmonėms, kurie nėra DI ekspertai, šiame dokumente pateikta informacija taip pat bus naudinga kaip pradinė mokomoji priemonė, su kuria susipažinus vėliau bus galima šios mokslo ir technologijų srities žinias gilinti ir pasidomėti DI išsamiau bei nuodugniau.

Aukšto lygio ekspertų grupė DI klausimais yra Europos Komisijos 2018 m. birželio mėn. sudaryta nepriklausomų ekspertų grupė.

Kontaktinis asmuo: Nathalie Smuha – Aukšto lygio ekspertų grupės DI klausimais koordinatork  
E. paštas: CNECT-HLG-AI@ec.europa.eu

Europos Komisija  
B-1049 Brussels

Dokumentas paskelbtas 2019 m. balandžio X d.

**Pirmasis šio dokumento projektas paskelbtas 2018 m. gruodžio 18 d. kartu su Aukšto lygio ekspertų grupės DI klausimais parengtu patikimo DI etikos gairių projektu. Atsižvelgiant į pastabas, gautas per Europos DI aljansą ir viešas konsultacijas dėl gairių projekto, dokumentas buvo pakoreguotas. Norime ypač nuoširdžiai padėkoti visiems, kurie teikė savo nuomonę dėl pirmojo dokumento projekto.**

Nei Europos Komisija, nei kiti Komisijos vardu veikiantys asmenys neatsako už tai, kaip ši informacija gali būti naudojama. Už šio darbinio dokumento turinį yra atsakinga tik Aukšto lygio ekspertų grupė DI klausimais. Nors Komisijos tarnybos padėjo rengti šį dokumentą, jame išdėstyti teiginiai atspindi Aukšto lygio ekspertų grupės DI klausimais nuomonę ir jokiais aplinkybėmis negali būti laikomi oficialia Europos Komisijos pozicija.

Daugiau informacijos apie Aukšto lygio ekspertų grupę DI klausimais galima rasti internete (<https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>).

Europos Komisijos dokumentų pakartotinis panaudojimas reglamentuojamas Sprendimu 2011/833/ES (OL L 330, 2011 12 14, p. 39). Naudoti ar atgaminti nuotraukas ir kitą medžiagą, kurios autorių teisės nepriklauso Europos Sąjungai, galima tik gavus autorių teisių subjektų leidimą.

# DI APIBRĖŽTIS.

## PAGRINDINIAI PAJĖGUMAI IR MOKSLO ŠAKOS

Pirmiausia remiamės Europos Komisijos komunikate dėl dirbtinio intelekto (DI)<sup>1</sup> pasiūlyta DI apibrėžtimi:

*„Dirbtinis intelektas – tai sistemos, kurios elgiasi protingai, analizuodamos savo aplinką ir darydamos gana savarankiškus sprendimus tikslui pasiekti.*

*Dirbtinio intelekto sistemos gali būti grindžiamos vien tik programine įranga ir veikti virtualiajame pasaulyje (pvz., balso sintezatoriai, vaizdo analizės programinė įranga, paieškos sistemos, kalbos ir veido atpažinimo sistemos) arba gali būti integruotos techninėje įrangoje (pvz., pažanguose robotuose, savaeigėse transporto priemonėse, bepiločiuose orlaiviuose ar daiktų interneto objektuose)“.*

Savo dokumente šią apibrėžtį išplečiame ir aiškiau išdėstome tam tikrus DI, kaip mokslo šakos ir technologijos, aspektus. Taip norime užtikrinti, kad nekiltų nesusipratimų, kad visų bendros žinios apie DI būtų tokios pačios ir tas žinias net ir ne DI ekspertai galėtų veiksmingai panaudoti. Taip pat norime pateikti naudingos informacijos, kuria būtų galima remtis aptariant ir DI etikos gaires, ir DI politikos rekomendacijas.

### 1. DI sistemos

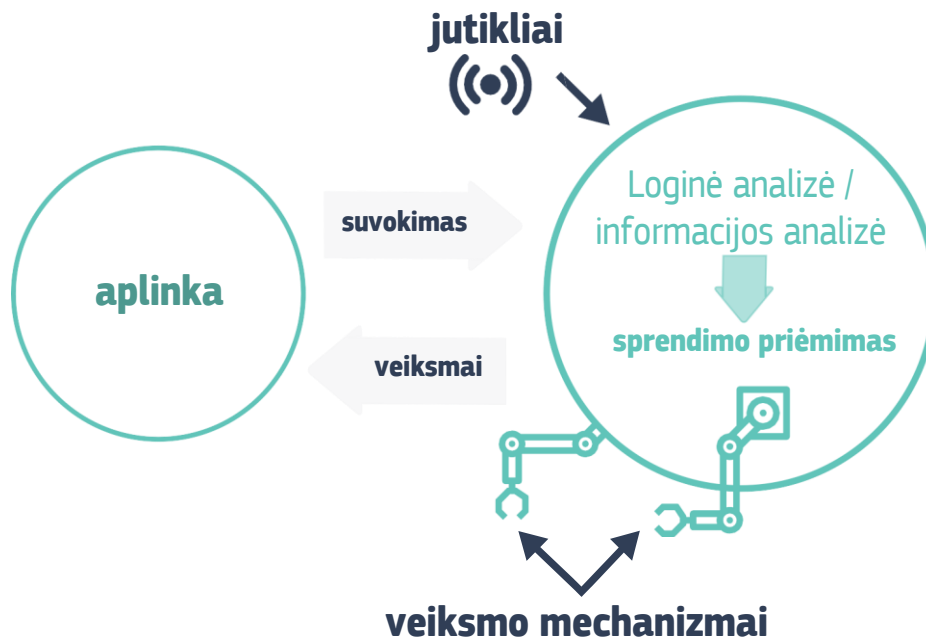
DI sąvokoje aiškiai minima intelekto sąvoka. Nors intelektą nuodugniai tyrinėja psichologai, biologai ir neuromokslininkai, intelektas (ir mašinų, ir žmonių) yra aptaki sąvoka, tad DI tiriantys mokslininkai daugiausia vartoja racionalumo sąvoką. Racionalumas – tai gebėjimas pasirinkti tinkamiausią veiksmą tam tikram tikslui pasiekti, remiantis tam tikrais optimizavimo kriterijais ir naudojantis turimais ištekliais. Žinoma, DI koncepcija susideda ne vien iš racionalumo, tačiau racionalumas yra reikšminga šios koncepcijos dalis.

Toliau kalbėdami apie komponentus, kuriuose naudojamas DI (ir apie programinę, ir apie aparatinę įrangą), vartosime sąvoką „DI sistema“. Iš tiesų, paprastai DI sistemos būna *integruotos* į didesnes sistemas kaip jų sudedamosios dalys, o ne savarankiškos sistemos.

Taigi, remiantis vienu iš populiariausių DI vadovėlių<sup>2</sup>, DI sistema, visų pirma, yra racionali. Tačiau koku būdu DI sistema veikia racionaliai? Kaip paminėta pirmiau pateiktos darbinės DI apibrėžties pirmame sakinyje, DI sistema tam tikrais savo jutikliais analizuoja aplinką, kurioje yra, taip renka ir aiškina duomenis, logiškai analizuoja tai, ką užfiksavo, arba apdoroja pagal tuos duomenis gautą informaciją, o tada nusprendžia, kokius veiksmus geriausia atlikti, ir tam tikrais veiksmo mechanizmais juos atlieka, tuo pačiu galbūt keisdama aplinką. DI sistemos gali naudoti simbolines taisykles arba sudaryti skaitmeninį modelį, jos taip pat gali koreguoti savo elgesį analizuodamos, kokį poveikį aplinkai padarė jų ankstesni veiksmas. 1 pav. pateiktas DI pavyzdys galbūt padės tai suprasti geriau.

<sup>1</sup> Komisijos komunikatas Europos Parlamentui, Europos Vadovų Tarybai, Europos ekonomikos ir socialinių reikalų komitetui ir Regionų komitetui „Dirbtinis intelektas Europai“, Briuselis, 2018 4 25, COM(2018) 237 final.

<sup>2</sup> Russell, S ir Norvig, P. „Artificial Intelligence: A Modern Approach“, Prentice Hall, 3rd edition, 2009.



1 pav. DI sistemos schema

**Jutikliai ir suvokimas.** 1 pav. sistemos jutikliai pavaizduoti WIFI simboliu. Iš tikrųjų tai gali būti kameros, mikrofonai, klaviatūra, interneto svetainė arba kiti įvesties įrenginiai, taip pat fizinių sąlygų (pvz. temperatūros, slėgio, atstumo, jėgos, traukos, taktiliniai) jutikliai. Iš esmės DI sistemoje reikia įrengti jutiklius, galinčius fiksuoti aplinkos duomenis, kurie yra reikalingi DI sistemą suprojektavusio žmogaus tai sistemai nustatytam tikslui pasiekti. Pavyzdžiui, jeigu norime sukurti DI sistemą, kuri automatiškai valytų kambario grindis, kai ant jų susikaupia purvo, vieni iš jutiklių galėtų būti grindų vaizdą fiksuojančios kameros.

Renkamus duomenis dažnai yra naudinga suskirstyti į struktūruotus ir nestrukūruotus. *Struktūruoti duomenys* – tai pagal apibrėžtą modelį surūšiuoti duomenys (pavyzdžiui, tarpusavio ryšių duomenų bazė), o *nestrukūruoti duomenys* nėra niekaip sąmoningai surūšiuoti (pavyzdžiui, nuotrauka arba teksto dalis).

**Loginė analizė arba informacijos apdorojimas ir sprendimų priėmimas.** DI sistemos esmę sudaro jos loginės analizės arba informacijos apdorojimo modulis, kuris naudoja iš jutiklių gautus įvesties duomenis ir, remdamasis nustatytu tikslu, pasiūlo, kokį veiksmą atlikti. Tai reiškia, kad jutiklių surinktus duomenis reikia paversti informacija, kurią loginės analizės arba informacijos apdorojimo modulis gali suprasti. Pavyzdyje paminėtoje valymo DI sistemoje kamera užfiksuoja grindų vaizdą ir jį perduos loginės analizės arba informacijos apdorojimo moduliui, o šis turės nuspręsti, ar grindis reikia valyti, ar ne (tai yra, koks veiksmas yra tinkamiausias norimam tikslui pasiekti). Nors mums, žmonėms, gali atrodyti lengva pažiūrėjus į nuotrauką nuspręsti, ar grindis reikia valyti, mašinai tai padaryti ne taip paprasta, nes nuotrauka jai tėra 0 ir 1 seka. Todėl loginės analizės arba informacijos apdorojimo modulis turi:

1. suprasti nuotrauką, kad galėtų nuspręsti, ar grindys yra švarios, ar ne. Iš esmės tai reiškia, kad sistema turi gebėti duomenis paversti informacija ir ją greitai modeliuoti, tačiau informacija turi apimti visus svarbius duomenis (šiuo atveju – ar grindys yra švarios, ar ne);
2. šias žinias logiškai analizuoti arba šią informaciją apdoroti ir sudaryti skaitmeninį modelį (tai yra, matematinę formulę), tada pagal ją nuspręsti, kokį veiksmą geriausia atlikti. Pavyzdžiui, jeigu iš nuotraukos gaunama informacija, kad grindys yra purvinos, tinkamiausias veiksmas yra aktyvuoti valymo funkciją, kitu atveju tinkamiausias veiksmas – nieko nedaryti.

Atkreipkite dėmesį, kad sąvoką „sprendimas“ reikėtų suprasti plačiaja prasme, kaip bet kurį veiksmą renkantis, ką daryti; tai nebūtinai reiškia, kad DI sistemos yra visiškai autonomiškos. Sprendimas taip pat gali būti ir rekomendacijų rinkinys, pateikiamas žmogui, kad žmogus priimtų galutinį sprendimą.

**Veiksmo įjungimas.** Kai DI sistema nusprendžia, kokį veiksmą atlikti, ji būna pasirengusi tą veiksmą atlikti joje sumontuotais veiksmo mechanizmais. Pirmiau pateiktame paveikslėlyje veiksmo mechanizmai pavaizduoti kaip rankos, tačiau jie nebūtinai turi būti fiziniai. Veiksmo mechanizmų funkciją gali atlikti ir programinė įranga. Jeigu pateiktame valymo pavyzdyje reikalingas veiksmas yra išsiurbti grindis, DI sistema gali pasiųsti signalą, kuriuo bus įjungtas dulkių siurblys. Kitas pavyzdys – pokalbių sistema (tai yra, pokalbių robotas), kuri renka tekstą ir taip atsako į naudotojo frazes.

Atliktas veiksmas greičiausiai pakeis aplinką, todėl kitą kartą sistemai reikės vėl pasinaudoti savo jutikliais ir iš pakitusios aplinkos surinkti galbūt pakitusią informaciją.

Racionalios DI sistemos ne visada kaip savo tikslą pasirenka tinkamiausią veiksmą, tad jos *racionalumas yra ribotas* dėl išteklių, pavyzdžiui, laiko arba skaičiavimo galios, ribotumo.

*Racionalios DI sistemos* yra pati paprasčiausia DI sistemos versija. Jos keičia aplinką, tačiau savo elgsenos laikui bėgant nekeičia, kad tikslą pasiektų efektyviau. *Besimokanti racionali sistema* yra racionali sistema, kuri, atlikusi veiksmą, įvertina naują aplinkos padėtį (ją analizuodama) ir nustato, kiek jos veiksmas buvo sėkmingas, o tada koreguoja savo loginės analizės taisykles ir sprendimų priėmimo metodus.

## 2. DI kaip mokslo šaka

Pirmiau pateiktas labai supaprastintas ir abstraktus DI sistemos apibūdinimas išskiriant tris pajėgumus: suvokimo, loginės analizės arba sprendimų priėmimo ir aktyvavimo. Tačiau toks apibūdinimas pakankamas, kad galėtume pristatyti ir suprasti daugumą DI metodų ir mokslo šakos sričių, kurios šiuo metu yra naudojamos kuriant DI sistemas, nes jos visos paremtos įvairiais sistemų pajėgumais. Bendrai tariant, visus metodus galima suskirstyti į dvi pagrindines grupes pagal pajėgumą atlikti *loginę analizę* ir *mokytis*. Dar viena svarbi mokslo šaka yra robotika.

**Loginė analizė ir sprendimų priėmimas.** Prie šios metodų grupės priskirti metodai – žinių vaizdavimas ir loginė analizė, veiksmų ir laiko planavimas, paieška ir optimizavimas. Šie metodai leidžia atlikti jutikliais surinktų duomenų loginę analizę. Kad tai būtų galima padaryti, duomenis reikia paversti žiniomis, tad viena DI sritis tas žinias turi kuo geriau sumodeliuoti (*žinių vaizdavimas*). Žinias sumodeliavus, reikia pereiti prie kito etapo – naudojantis tomis žiniomis, atlikti loginę analizę (*žinių loginę analizę*). Per šį procesą taikomos simbolinės taisyklės, *planuojami veiksmai ir laikas*, iš daugybės variantų *ieškoma* sprendimo, iš visų galimų sprendimų pasirenkamas *optimalus* problemos sprendimo būdas. Paskutiniu etapu reikia nuspręsti, kokį veiksmą atlikti. DI sistemos dalį sudaranti loginė analizė ir sprendimų priėmimas dažnai yra labai sudėtingi procesai, per juos reikia derinti keletą pirmiau paminėtų metodų.

**Mokymasis.** Prie šios metodų grupės priskiriamas mašinų mokymasis, neuroniniai tinklai, gilusis mokymasis, sprendimų medžiai ir daug kitų mokymosi metodų. Šie metodai leidžia DI sistemai mokytis, kaip spręsti problemas, kurių neįmanoma tiksliai apibrėžti arba kurių sprendimo neįmanoma apibrėžti pagal simbolines loginės analizės taisykles. Pavyzdžiui, tokios problemos yra problemos, susijusios su suvokimo pajėgumais, pavyzdžiui, *kalba, kalbos supratimas, kompiuterio matomas vaizdas, elgesio prognozė*. Šios problemos atrodo paprastos, nes žmonėms jas išspręsti ištis lengva. Tačiau DI sistemoms jos nėra tokios paprastos, nes DI sistemos atlikdamos loginę analizę negali (bent jau kol kas) kliautis sveiku protu. Šios problemos tampa ypač sudėtingomis, kai sistemai tenka suprasti nestruktūruotus duomenis. Šiuo atveju padeda pagal *mašinų mokymosi* principus veikiantys metodai. Tačiau mašinų mokymosi metodus galima naudoti daug įvairesnėms užduotims vykdyti, ne vien suvokimo. Pagal mašinų mokymosi metodus sudaromas skaitmeninis modelis (tai yra, matematinė formulė), kurią pritaikius duomenims, apskaičiuojamas sprendimas.

Mašinų mokymasis vyksta keletu būdų. Plačiausiai paplitę būdai yra *prižiūrimas mokymasis, neprižiūrimas mokymasis* ir *sustiprintas mokymasis*.

Pagal prižiūrimo mašinų mokymosi metodą užuot nustačius sistemos elgesio taisykles, jai pateikiami pradinio ir baigtinio elgesio pavyzdžiai tikintis, kad sistema sugebės apibendrinti šiuos pavyzdžius (dažniausiai tai, kas jau įvyko) ir gebės tinkamai elgtis, kai iškils pavyzdžiuose neparodyti atvejai (jie gali iškilti ateityje). Pateiktame pavyzdyje galėjome pateikti sistemai daug grindų nuotraukų su atitinkamu paaiškinimu (tai yra, ar nuotraukoje pavaizduotos

grindys yra švarios, ar ne). Jeigu pateiksime pakankamai pavyzdžiui, kurie bus pakankamai įvairūs ir pakankamai įtraukiai apims daugumą galimų atvejų, naudodamasi mašinų mokymosi algoritmu sistema galės apibendrinti žinias, kaip tinkamai suprasti anksčiau nematytų grindų vaizdą. Pagal tam tikrą mašinų mokymosi metodiką yra naudojami algoritmai, sukurti vadovaujantis *neuroninių tinklų* koncepcija, kuri bendrai yra įkvėpta to, kaip veikia žmogaus smegenys, tai yra, neuroninius tinklus sudaro mažų tarpusavyje daugybe jungčių, kurioms priskirtos svorinės vertės, susietų procesorių (analogiškų mūsų neuronams) tinklas. Į neuroninį tinklą pateikiami įvesties duomenys yra jutikliais surinkti duomenys (pateiktame pavyzdyje – grindų nuotrauka), o gauta išvestis – nuotraukos supratimas (mūsų pavyzdyje – ar grindys švarios, ar ne). Kai pavyzdžiai analizuojami (vyksta tinklo *mokymasis*), jungčių svorinės vertės yra koreguojamos, kad kuo geriau atitiktų turimus pavyzdžius (tai yra, siekiant sumažinti klaidas tarp numatomo rezultato ir tinklo apskaičiuoto rezultato). Mokymosi etapo pabaigoje vyksta neuroninio tinklo elgesio bandymas naudojant prieš tai nematytus pavyzdžius, taip siekiant patikrinti, ar užduotis išmokta gerai.

Svarbu pažymėti, kad šiai metodikai (kaip visiems mašinų mokymosi metodams) būdingas tam tikslas procentinis klaidų lygis, tačiau šis lygis paprastai būna labai mažas. Taigi, svarbiausias rodiklis yra *tikslumas* – matas, parodantis, kokia yra procentinė teisingų atsakymų dalis.

Yra kelios neuroninių tinklų rūšys ir keletas mašinų mokymosi metodikų. Šiuo metu sėkmingiausia iš jų yra *gilusis mokymasis*. Ši metodika paremta tuo, kad neuroniniuose tinkluose įvestį ir išvestį skiria keletas sluoksnių, leidžiančių laipsniškai suprasti bendrą įvesties ir išvesties ryšį. Dėl to bendra metodika yra tikslesnė, pagal ją reikalingas ne toks intensyvus žmogaus vadovavimas.

Neuroniniai tinklai tėra viena mašinų mokymosi priemonė, tačiau tokių priemonių yra daug daugiau ir jos pasižymi įvairiais ypatumais. Tai: „atsitiktinių miškų“ (angl. *random forests*) ir „patobulintų medžių“ (angl. *boosted trees*), klasterių sudarymo (angl. *clustering*), matricos suskaidymo (angl. *matrix factorization*) metodika ir pan.

Dar viena naudinga mašinų mokymosi metodika vadinama *sustiprintu mokymusi*. Pagal šią metodiką DI sistema tam tikrą laiką gali laisvai priimti sprendimus. Kiekvieną kartą, sistemai priėmus sprendimą, jai pasiunčiamas grįžtamasis signalas, kuriuo pažymima, ar sprendimas buvo geras, ar blogas. Sistemos tikslas – ilgainiui gauti kuo daugiau teigiamų grįžtamųjų signalų. Pavyzdžiui, tokia metodika naudojama rekomendacijų sistemoje (pavyzdžiui, keliose internete veikiančiose rekomendacijų sistemose, kurios siūlo naudotojams jiems galinčius patikti pirkinys) arba rinkodaros srityje.

Mašinų mokymosi metodikos yra naudingos ne tik atliekant suvokimo užduotis, pavyzdžiui, vaizdo ir teksto suvokimo užduotis, bet ir kitas užduotis, kurias sunku apibrėžti ir neįmanoma išsamiai aprašyti simbolinėmis elgesio taisyklėmis.

Atkreipkite dėmesį į skirtumą tarp mašinų mokymosi metodikų, pagal kurias mokomasi naujos užduoties, kurios neįmanoma aprašyti simboliu būdu, ir racionalių subjektų mokymosi (jis paminėtas ankstesniame skirsnyje), kai ilgainiui elgesys yra koreguojamas, kad nustatytas tikslas būtų pasiektas efektyviau. Šios dvi metodikos gali dubliuotis arba papildyti viena kitą, tačiau nebūtinai sutampa.

**Robotika.** Robotiką galima apibrėžti kaip „DI veikimą fiziniame pasaulyje“ (ji dar vadinama *integruotu DI*). Robotas yra fizinė mašina, kuri turi įveikti fizinio pasaulio dinamiką, netikrumą ir sudėtingumą. Suvokimas, loginė analizė, veiksmai, mokymasis, sąveikos su kitomis sistemomis galimybės yra dažniausiai integruojamos į roboto sistemos kontrolės architektūrą. Projektuojant robotą ir jam veikiant svarbus ne tik DI, bet ir kitos mokslo šakos, pavyzdžiui, mechaninė inžinerija ir kontrolės teorija. Robotų pavyzdžiai – robotai manipulatoriai, autonominės transporto priemonės (pvz., automobiliai, bepilotės skraidyklės, skraidantys taksi), robotai humanoidai, robotai dulkių siurbliai ir pan.

2 pav. pavaizduota dauguma pirmiau paminėtų DI mokslo šakų sričių ir jų tarpusavio ryšiai. Tačiau svarbu paminėti, kad DI yra daug sudėtingesnis negu pavaizduota paveikslėlyje, nes jis apima daug kitų mokslo šakų sričių ir metodų. Be to, kaip jau buvo paminėta, robotikoje naudojami ir kiti DI sričiai nepriklausantys metodai. Tačiau tikime, jog to pakanka, kad būtų galima aiškiai suprasti, kokius informacijos mainų, sklaidos, DI, DI etikos ir DI politikos klausimus reikia aptarti daugiašakėje ir daug suinteresuotųjų subjektų vienijančioje aukšto lygio ekspertų grupėje.



2 pav. Supaprastinta DI mokslo šakų sričių ir jų tarpusavio ryšių apžvalga.

Ir mašinių mokymuisi, ir loginėi analizei naudojama daug kitų metodų. Robotikoje naudojami metodai, kurie nėra susiję su DI. Visa DI sritis priskirtina kompiuterių mokslo šakai.

### 3. Kitos svarbios DI sąvokos ir klausimai

**Siaurasis (arba silpnasis) ir bendrasis (arba stiprusis) DI.** Bendroji DI sistema – tai sistema, kuri gali atlikti daugumą žmonių atliekamų veiksmų. Siaurosios DI sistemos – tai sistemos, kurios gali atlikti vieną arba kelias konkrečias užduotis. Šiuo metu diegiamos DI sistemos yra siaurojo DI pavyzdžiai. DI kūrimo pradžioje tyrėjai vartojo kitokias sąvokas (silpnasis ir stiprusis DI). Kuriant reikalingus bendrojo DI pajėgumus, dar reikia išspręsti daug neišspręstų etikos, mokslo ir techninių problemų, pavyzdžiui sveiku protu pagrįstos loginės analizės, savimonės, mašinos gebėjimo apibrėžti savo pačios tikslą.

**Duomenų klausimai ir šališkumas.** Kadangi tam, kad užduotys būtų atliekamos tinkamai, daugelis DI sistemų, pavyzdžiui, sistemos, į kurias įdiegti prižiūrimo mašinių mokymosi komponentai, naudojami didžiuliu kiekiu duomenų, svarbu suprasti, kaip duomenys veikia DI sistemos elgesį. Pavyzdžiui, jeigu mokymui skirti duomenys yra šališki, tai yra, nėra pakankamai objektyvūs arba įtraukūs, pagal tokius duomenis mokoma DI sistema negalės daryti tinkamų apibendrinančių išvadų ir greičiausiai priims neteisingus sprendimus, kurie vienoms grupėms bus palankesni negu kitoms. Pastaruoju metu DI bendruomenė kuria metodus, skirtus nustatyti ir mažinti šališkumą mokymui skirtų duomenų rinkiniuose ir kitose DI sistemos dalyse.

**Juodosios dėžės DI ir paaiškinamumas.** Kai kurie mokymosi metodai tikslumo požiūriu yra labai sėkmingi, tačiau yra labai neaiškūs norint suprasti, kaip pagal juos priimami sprendimai. *Juodosios dėžės DI* – tai tokie atvejai, kai tam tikrų sprendimų logiškai paaiškinti neįmanoma. DI sistemų paaiškinamumas pasireiškia tuo, kad vietoj to jos gali tam tikra forma paaiškinti savo veiksmus.

**Į tikslą orientuotas DI.** Dabartinės DI sistemos yra orientuotos į tikslą, tai yra, žmogus joms pateikia tikslo, kurį reikia pasiekti, specifikaciją ir jos pasinaudoja tam tikrais metodais, kad tą tikslą pasiektų. Tačiau savų tikslų jos

neužsibrėžia. Tačiau kai kuriose DI sistemose (pavyzdžiui, DI sistemose, kuriose naudojami tam tikri mašinų mokymosi metodai) suteikiama daugiau laisvės spręsti, kurį būdą pasirinkti, kad nustatytas tikslas būtų pasiektas.

#### 4. Papildyta DI apibrėžtis

Siūlome naudoti šią papildytą DI apibrėžtį:

„Dirbtinio intelekto (DI) sistemos yra žmonių sukurtos<sup>3</sup> programinės įrangos (taip pat gali būti aparatinės įrangos) sistemos, kurios, joms nustačius sudėtingą tikslą, veikia fiziniu ir skaitmeniniu lygmenimis – analizuoja savo aplinką rinkdamos duomenis, aiškina surinktus struktūruotus ir nestruktūruotus duomenis, logiškai analizuoja turimas žinias arba apdoroja pagal tuos duomenis suformuotą informaciją ir priima sprendimą, kokį (-ius) veiksmą (-us) geriausia atlikti, kad užsibrėžtas tikslas būtų pasiektas. DI sistemos gali naudoti simbolines taisykles arba sudaryti skaitmeninį modelį, jos taip pat gali koreguoti savo elgesį analizuodamos, kokį poveikį aplinkai padarė jų ankstesni veiksmai.

Kaip mokslo šaka, DI apima keletą metodikų ir metodų, pavyzdžiui, mašinų mokymąsi (konkretūs pavyzdžiai – gilusis mokymasis ir sustiprintas mokymasis), mašinų atliekamą loginę analizę (įskaitant veiksmų ir laiko planavimą, žinių vaizdavimą, loginę analizę, paiešką ir optimizavimą) ir robotiką (įskaitant kontrolę, suvokimą, jutiklius, veiksmo mechanizmus ir visų kitų metodų integravimą į kibernetines - fizines sistemas)“.

Šiai apibrėžčiai pagrįsti kaip papildomos informacijos šaltiniu siūlome remtis šiuo dokumentu.

---

<sup>3</sup> Žmonės tiesiogiai projektuoja DI sistemas, tačiau jie taip pat gali panaudoti DI metodus savo projektams optimizuoti.



## Šį dokumentą parengė aukšto lygio ekspertų grupės DI klausimais nariai

Nariai išvardyti abėcėlės tvarka.

Pekka Ala-Pietilä, Aukšto lygio ekspertų grupės DI klausimais pirmininkė „Al Finland“, „Huhtamaki“, „Sanoma“	Pierre Lucas „Orgalim“ – Europos technologijų sektoriaus organizacija
Wilhelm Bauer „Fraunhofer“	Ieva Martinkėnaitė „Telenor“
Urs Bergmann „Zalando“	Thomas Metzinger J. Gutenbergo universitetas (Maincas) ir Europos universitetų asociacija
Mária Bielíková Slovakijos technologijų universitetas, Bratislava	Catelijne Muller „ALLAI Netherlands“ ir EESRK
Cecilia Bonefeld-Dahl „DigitalEurope“	Markus Noga SAP
Yann Bonnet ANSSI	Barry O’Sullivan, Aukšto lygio ekspertų grupės DI klausimais pirmininko pavaduotojas Korko universitetinis koledžas
Loubna Bouarfa OKRA	Ursula Pacht BEUC
Stéphan Brunessaux „Airbus“	Nicolas Petit Liežo universitetas
Raja Chatila IEEE iniciatyva „Intelektinių / autonominių sistemų etika“ (angl. <i>Ethics of Intelligent/Autonomous Systems</i> ) ir Sorbonos universitetas	Christoph Peylo „Bosch“
Mark Coeckelbergh Vienos universitetas	Iris Plöger BDI
Virginia Dignum Umėjos University	Stefano Quintarelli „Garden Ventures“
Luciano Floridi Oksfordo universitetas	Andrea Renda Europos koledžo fakultetas ir CEPS
Jean-Francois Gagné „Element AI“	Francesca Rossi* IBM
Chiara Giovannini ANEC	Cristina San José Europos bankų federacija
Joanna Goodey Pagrindinių teisių agentūra	George Sharkov „Digital SME Alliance“
Sami Haddadin Miuncheno robotikos ir MI mokykla	Philipp Slusallek Vokietijos DI tyrimų centras (DFKI)
Gry Hasselbalch Mokslinių tyrimų ir veiklos centras „DataEthics“ ir Kopenhagos universitetas	Françoise Soulié Fogelman Konsultantas DI klausimais
Fredrik Heintz Linkopingo universitetas	Saskia Steinacker „Bayer“
Fanny Hidvegi „Access Now“	Jaan Tallinn „Abient Sound Investment“
Eric Hilgendorf Vurcburgo universitetas	Thierry Tingaud „STMicroelectronics“
Klaus Höckner „Hilfsgemeinschaft der Blinden und Sehschwachen“	Jakob Uszkoreit „Google“
Mari-Noëlle Jégo-Laveissière „Orange“	Aimee Van Wynsberghe „TU Delft“
Leo Kärkkäinen „Nokia Bell Labs“	Thiébaud Weber ETUC
Sabine Theresia Köszegi „TU Wien“	Cecile Wendling AXA
Robert Kroplewski Lenkijos vyriausybės atstovas ir patarėjas	Karen Yeung Birmingemo universitetas
Elisabeth Ling RELX	

\*Francesca Rossi yra šio dokumento pranešėja.