

Etiniai aspektai ir gairės

Parengė: dr. Aušra Vidugirienė

Pagal: **Dirbtinio intelekto (DI) ir duomenų naudojimo mokymo
ir mokymosi srityje etikos gairės pedagogams**

Patikimam DI keliami pagrindiniai reikalavimai

- **žmogiškasis veiksnys ir žmogaus atliekama priežiūra**, apimantys pagrindines teises, vaikų teises, žmogiškąjį veiksni ir žmogaus atliekamą priežiūrą;
- **skaidrumas**, apimantis atsekamumą, paaiškinamumą ir informacijos sklaidą;
- **įvairovė, nediskriminavimas ir teisingumas**, apimantys prieinamumą, universalumą, nepagrįsto šališkumo vengimą ir suinteresuotųjų subjektų dalyvavimą, dėl kurio sistemomis galima naudotis nepaisant amžiaus, lyties, gebėjimų ar savybių, ypatingą dėmesį skiriant specialiųjų poreikių turintiems besimokantiems asmenims;
- **visuomenės ir aplinkos gerovė**, apimanti tvarumą ir palankumą aplinkai, socialinį poveikį, visuomenę ir demokratiją;
- **privatumas ir duomenų valdymas**, apimantys pagarbą privatumui, duomenų kokybę ir vientisumą, taip pat prieigą prie duomenų;
- **techninis patvarumas ir saugumas**, apimantys atsparumą išpuoliams, saugumą ir bendrąją saugą, tikslumą, patikimumą ir atkuriamumą;
- **atskaitomybė**, apimanti galimybę atlikti auditą, neigiamo poveikio sumažinimą ir informavimą apie jį, kompromisus ir teisių gynimą.

Žmogiškasis veiksnys ir žmogaus atliekama priežiūra

- Ar mokytojo vaidmuo aiškiai apibrėžtas užtikrinant, kad naudojant DI sistemą cikle dalyvautų ir mokytojas? Kaip DI sistema veikia mokytojo mokomąjį vaidmenį?
- Ar besimokantiems asmenims poveikį darantys sprendimai įgyvendinami veikiant mokytojui ir ar mokytojas gali pastebėti nenormalią situaciją ar galimą diskriminaciją?
- Ar įdiegtos procedūros, kad mokytojai galėtų stebėti ir įsikišti, pvz., situacijose, kuriose bendraujant su besimokančiais asmenimis ar tėvais reikia empatijos?
- Ar yra mechanizmas, leidžiantis besimokantiems asmenims atsisakyti dalyvauti, jeigu į susirūpinimą keliančius dalykus tinkamai nereaguoja?
- Ar įdiegtos stebėsenos sistemos, kad DI sistema nebūtų pernelyg pasitikima ir pasikliaujama?
- Ar mokytojai ir mokyklos vadovai yra gavę visus reikiamus mokymus ir informaciją veiksmingai naudotis sistema ir užtikrinti, kad ji būtų saugi ir nedarytų žalos ar nepažeistų besimokančių asmenų teisių?

Skaidrumas

- Ar mokytojai ir mokyklos vadovai žino, kokie DI metodai ir funkcijos naudojami sistemoje?
- Ar aišku, kuriuos aspektus DI sistemoje gali perimti, o kurių ne?
- Ar mokytojai ir mokyklos vadovai supranta, kaip DI sistemoje veikia konkretūs vertinimo ar individualizavimo algoritmai?
- Ar sistemos procesai ir rezultatai orientuoti į numatomus besimokančių asmenų mokymosi rezultatus? Kiek DI sistemos predikcijos, vertinimai ir klasifikacijos yra patikimi aiškinant ir vertinant jos naudojimo aktualumą?
- Ar instrukcijos bei informacija prieinamos ir pateikiamos aiškiai tiek mokytojams, tiek besimokantiems asmenims?

Įvairovė, nediskriminavimas ir teisingumas

- Ar sistema visi gali naudotis vienodai be kliūčių?
- Ar sistema užtikrinami sąveikos būdai, tinkami negalia ar specialiųjų ugdymosi poreikių turintiems besimokantiems asmenims? Ar DI sistema sukurta taip, kad su besimokančiais asmenimis būtų elgiamasi pagarbiai, atsižvelgiant į jų individualius poreikius?
- Ar naudotojo sąsaja tinkama ir prieinama pagal besimokančių asmenų amžių? Ar tinkamumas tikslinei amžiaus grupei ir tos amžiaus grupės naudotojų patirtis išbandyti?
- Ar įdiegtos procedūros siekiant užtikrinti, kad naudojant DI jokie naudotojai nepatirtų diskriminacijos ar su jais nebūtų elgiamasi neteisingai?
- Ar iš DI sistemos dokumentų arba jos mokymo proceso galima įžvelgti galimą duomenų šališkumą?
- Ar įdiegtos procedūros galimam šališkumui ar suvokiamai nelygybei nustatyti ir mažinti

Visuomenės ir aplinkos gerovė

- Kaip DI sistema veikia socialinę ir emocinę besimokančių asmenų ir mokytojų gerovę?
- Ar DI sistema aiškiai nurodo, kad imituoja socialinę sąveiką ir neturi gebėjimo jausti ar empatijos?
- Ar besimokantys asmenys arba jų tėvai dalyvavo priimant sprendimą naudoti DI sistemą ir tam pritarė?
- Ar duomenys naudojami siekiant padėti mokytojams ir mokyklos vadovams įvertinti besimokančių asmenų gerovę; jei taip, kaip tai stebima?
- Ar naudojant sistemą asmenims ar visuomenei daroma kokia nors žala arba keliama baimė?

Privatumas ir duomenų valdymas

- Ar yra mechanizmų, skirtų užtikrinti, kad neskelbtini duomenys liktų anoniminiai? Ar įdiegtos procedūros, kad duomenis galėtų gauti tik tie asmenys, kuriems jų reikia?
- Ar prieiga prie besimokančių asmenų duomenų saugoma, užtikrinama saugioje vietoje ir naudojama tik tiems tikslams, kuriems buvo renkami duomenys?
- Ar yra mechanizmas, kad mokytojai ir mokyklos vadovai galėtų pranešti apie problemas, susijusias su privatumu ar duomenų apsauga?
- Ar besimokantis asmenys ir mokytojai yra informuoti apie tai, kas daroma su jų duomenimis, kaip ir kokiais tikslais jie naudojami?
- Ar įmanoma prisitaikyti privatumo ir duomenų nustatymus?
- Ar DI sistema atitinka Bendrąjį duomenų apsaugos reglamentą?

Techninis patvarumas ir saugumas

- Ar užtikrinamas pakankamas saugumas, kad būtų apsisaugota nuo duomenų saugumo pažeidimų?
- Ar parengta strategija, kaip stebėti ir tikrinti, ar DI sistema atitinka tikslus, paskirtį ir numatytus taikymo būdus?
- Ar įdiegti tinkami duomenų rinkimo, saugojimo, tvarkymo, kiekio mažinimo ir naudojimo priežiūros mechanizmai?
- Ar esama informacijos, kuri patikintų besimokančius asmenis ir tėvus, kad sistema techniškai patvari ir saugi?

Atskaitomybė

- Kas atsakingas už nuolatinę DI sistemos teikiamų rezultatų stebėseną ir kaip tie rezultatai naudojami mokymui, mokymuisi ir vertinimui gerinti?
- Kaip vertinamas DI sistemos veiksmingumas ir poveikis ir kaip vertinant atsižvelgiama į pagrindines švietimo vertybes?
- Kas atsakingas ir atskaitingas už galutinius sprendimus dėl DI sistemos įsigijimo ir įdiegimo?
- Ar pasirašytas susitarimas dėl paslaugų lygio, kuriame aiškiai nurodytos palaikymo ir techninės priežiūros paslaugos ir veiksmai, kurių bus imamasi nurodytomis problemoms išspręsti?

- Rekomendacija dėl dirbtinio intelekto etikos
- [Rekomendacija dėl dirbtinio intelekto etikos - UNESCO Digital Library](#)

Etiniai iššūkiai

<https://di.lt/dirbtinis-intelektas-viskas-ka-reikia-zinoti-apie-dirbtini-intelektu/>

- **Darbo rinkos pokyčiai:** Stiprios dirbtinio intelekto sistemos gali pakeisti žmonių vykdomas darbo funkcijas, o tai gali sukelti didelius socialinius ir ekonominius pokyčius. Privalome apsvarstyti, kaip užtikrinti, kad dirbtinis intelektas prisidėtų prie darbuotojų gerovės ir užtikrintų mokymo galimybes naujoms kompetencijoms įgyti.
- **Diskriminacija ir nelygybė:** Dirbtinio intelekto algoritmai gali turėti tendenciją atspindėti žmogiškus iškraipymus ir nelygybę, kuri egzistuoja duomenyse arba visuomenėje. Tai gali padidinti diskriminaciją ir neteisingumą skirtingoms socialinėms grupėms.
- **Privatumo ir duomenų saugumas:** Stiprios dirbtinio intelekto sistemos dažnai reikalauja didelės duomenų bazės, kad galėtų mokytis ir tobulėti. Svarbu užtikrinti, kad asmeniniai duomenys būtų saugiai ir etiškai tvarkomi, kad būtų išvengta duomenų pažeidimų arba piktnaudžiavimo.
- **Atsakomybė ir paaiškinimai:** Stiprios dirbtinio intelekto sistemos dažnai yra labai sudėtingos ir sunkiai suprantamos. Tai kelia iššūkius dėl atsakomybės už jų veikimą ir sprendimus. Būtina užtikrinti, kad sistemos būtų atvirose ir aiškiose sąlygose, kad galėtų paaiškinti savo sprendimus ir veikimo principus.
- **Kontrolė ir saugumas:** Stiprios dirbtinio intelekto sistemos gali tapti labai galingos ir netgi pereiti į autonomines formas. Tai kelia iššūkių dėl sistemos kontroliavimo ir užtikrinimo, kad ji neperžengtų žmogaus kontrolės ir patektų į nepageidaujamas situacijas.
- **Etiniai sprendimai ir konfliktai:** Stiprios dirbtinio intelekto sistemos gali rasti sprendimus, kurie yra teisingi ir tinkami iš techninės perspektyvos, bet kurie kelia etinius konfliktus arba prieštarauja žmonių vertybėms ir interesams.
- **Priklausomybė ir priklausomybė:** Stiprios dirbtinio intelekto sistemos gali kurti įpročius ir priklausomybę vartotojams, pavyzdžiui, socialiniuose tinkluose arba prekių pirkimo procese. Tai gali turėti neigiamą poveikį žmonių sveikatai ir gerovei.