

Dirbtinio intelekto naudojimo Vilniaus universitete gairės

Gairių paskirtis

Šiomis dirbtinio intelekto naudojimo Vilniaus universitete gairėmis (toliau – Gairės) siekiama pateikti rekomendacijas dėl tinkamo dirbtinio intelekto (toliau – DI) įrankių pasitelkimo mokslinėje, mokymo(si) veikloje siekiant išvengti akademinės etikos, autorystės teisinių normų pažeidimų ir asmeninės ar konfidencialios informacijos atskleidimo.

Dirbtinis intelektas ir DI generatyviniai modeliai

Dirbtinis intelektas (angl. *artificial intelligence*) – tai technologijos sritis, kurioje kuriami kompiuteriniai algoritmai, galintys atlikti užduotis, įprastai reikalaujančias žmogaus intelekto. Jie apima, bet neapsiriboja, kalbos supratimu, mokymusi, problemų sprendimu, matymu ar sprendimų priėmimu. DI algoritmai mokomi naudojant didelius duomenų kiekius ir įvairias mokymo strategijas, kad galėtų atpažinti raštus, kalbėti, vertinti ir kt.

DI generatyviniai modeliai – tai DI modelių kategorija, kuri ne tik analizuoja ir interpretuoja duomenis, bet ir geba generuoti naujus, panašius į išmokus duomenis. Šie modeliai mokomi naudojant statistinius metodus, kaip antai dirbtinius neuroninius tinklus, kad suvoktų ir atkartotų duomenų struktūras. Abi šios sritys patenka į platų ir sparčiai besivystantį mokslinių tyrimų ir technologijų taikymo lauką. Žinia, kad DI ir generatyviniai modeliai nuolat tobulėja mokslininkams ir inžinieriams tobulinant mokymo metodus ir algoritmus.

Technologinė pažanga leidžia DI generatyvinius modelius integruoti į įvairias aplikacijas – nuo mobiliųjų programėlių iki kompleksinių mokslinių tyrimų sistemų. Dėl šios integracijos ne tik efektyviau veikia aplikacijos, bet ir kyla įvairių su saugumu ir etika susijusių klausimų.

DI generatyvinių modelių naudojimas studijų procese

DI generatyviniai modeliai gali praversti mokantis ir mokant universitete. Dėstytojams jie gali padėti taupyti laiką ruošiantis mokymo(si) veikloms (pvz., rengiant iliustracijas, lenteles, skaidres, praktines užduotis, generuojant idėjas, adaptuojant užduotis pagal individualius studentų poreikius ir pan.). Studentams jie gali padėti geriau išmokti (pvz., paaiškinti sudėtingas sąvokas, perprasti platesnį kontekstą, atsakyti į kylančius klausimus ir pan.) bei geriau ir greičiau atlikti užduotis (pvz., pagerinti rašymo stilių, teikti grįžtamąjį ryšį jau parengtoms užduotims, generuoti pirmines idėjas juodraščiui ir pan.). Tuo pat metu DI generatyviniai modeliai kelia tam tikrų iššūkių ir rizikų, susijusių su saugiu ir skaidriu duomenų panaudojimu, galimomis klaidomis ar šališkumais bei akademinė etika.

DI generatyviniai modeliai studijų procese paprastai naudojami šiais tikslais:

- 1) **individualizuojant mokymą(si)** – DI generatyviniai modeliai gali kurti pritaikytas mokymo(si) medžiagas, atsižvelgiant į besimokančiojo mokymosi stilių, lygį ir domėjimosi sritis.
- 2) **interaktyvioms užduotims ir simuliacijoms** – DI generatyviniai modeliai gali generuoti interaktyvias užduotis ir simuliacijas, kurios padeda besimokantiems geriau suprasti sudėtingas sąvokas.

Siekiant etiškai ir efektyviai naudoti tokio pobūdžio modelius ir programas, svarbu, kad visos suinteresuotos šalys savo veikloje laikytųsi toliau išdėstytų principų.

Naudotojų atsakomybės

1. Asmuo, skelbiantis ar naudojantis DI generatyvinio modelio sugeneruotus rezultatus, atsako už galutinio rezultato turinį ir kokybę. Atkreiptinas dėmesys, kad DI sugeneruotas turinys gali būti netikslus, klaidinantis arba neteisingas, pažeisti trečiųjų šalių intelektinės nuosavybės, asmens duomenų ar kitas saugomas teises. Dėl tos priežasties, bet koks pagalbiniais įrankiais sugeneruotas rezultatas turi būti peržiūrėtas ir patikrintas, o skaidrumas reikalauja atskleisti DI naudojimą.

2. DI generatyvinių modelių naudojimas tyrimams, rašto darbams, paraiškoms rengti ir pan. privalo būti aiškiai nurodytas (darbo įvade, aprašant tyrimo metodiką, arba pateikiant DI generatyvinių modelių naudojimo deklaraciją kaip darbo priedą). DI generatyvinių modelių naudojimas turi būti atskleidžiamas tyrimo grupei ir (arba) partneriams, nustatant aiškias jų naudojimo taisykles, formas, naudojamus duomenis ir kitus susijusius klausimus.

3. Rekomenduojama išsaugoti DI generatyvinių modelių sugeneruotus rezultatus visa apimtimi, o taip pat jų gavimo kelią, nurodyti naudotą DI generatyvinį modelį ir jo panaudojimo datą.

Administracijos atsakomybės:

1. užtikrinti aktualių dokumentų (KAP rašto metodiniai nurodymai, vertinimo kriterijai ir pan.) atnaujinimą, kuriuose, atsižvelgiant į studijų programų pobūdį ir poreikį, būtų apibrėžtos DI ir DI generatyvinių modelių naudojimo studijų procese normos ir atvejai, kai tokių įrankių naudojimas yra draudžiamas;

2. atsižvelgiant į studijų programų specifiką ir poreikius, parengti rekomendacijas deklaracijai dėl DI naudojimo rašto darbuose;

3. užtikrinti nuolatinę pagalbą ir paramą dėstytojams ir studentams DI klausimais, organizuojant periodiškus kompetencijų tobulinimo renginius;

4. teikti nuolatinę pagalbą studentams, dėstytojams bei Akademinės etikos komisijos nariams identifikuojant piktnaudžiavimo atvejus bei taikant atitinkamas sankcijas (pvz., plagiato patikros sistemų naudojimas, pagalba aptikti DI sugeneruotą tekstą pagal galimus požymius: a) tekstas neturi jokių nuorodų į šaltinius arba nurodyti šaltiniai neegzistuoja; b) tekste naudojama akivaizdžiai pasenusi informacija, daroma faktinių klaidų; c) tekstas parašytas neįprastai taisyklinga ir sklandžia kalba, tačiau analizė paviršutiniška, idėjos atsikartoja).

Dėstytojų atsakomybės:

1. atnaujinti kurso programas, ugdomas kompetencijas ir atsiskaitymo reikalavimus, atsižvelgiant į sparčią DI raidą. Rekomenduojama peržiūrėti iki šiol taikytas atsiskaitymo ir vertinimo strategijas, pabandant patikrinti, kaip DI programos geba atlikti kurso užduotis, jei reikia, jas adaptuoti, keisti (pvz., taikyti tam tikrų užduočių atsiskaitymo žodžiu arba „užverstos knygos“ egzamino strategijas, išdalinti užduotis laike ir sekti studentų progresą ir pan.);

2. su studentais aptarti etišką ir efektyvų DI generatyvinių programų naudojimą dalyko (modulio) mokymosi procese;

3. jei leidžiama naudoti DI generatyvinius modelius, tuomet atsakingai informuoti, pasidalinti ir diskutuoti su studentais apie tai, jog užduočių ar kitos medžiagos sudarymui buvo naudojamas DI;

4. kilus klausimams dėl etiško naudojimo, kreiptis į padalinio administraciją (Studijų skyriaus vadovą ar prodekaną studijoms) ar akademinės etikos komisiją;

5. nenaudoti DI generatyvinių modelių baigiamųjų darbų recenzijoms rengti. Suprantant, kad šios programos galėtų būti naudingos recenzuojant tam tikrus techninius baigiamųjų darbų aspektus (pvz., citavimą, žodžių skaičių, terminų vartojimą ar pan.), vis tik laikomasi pozicijos, kad klaidos rizika yra pernelyg didelė, ypač dėl to, kad dabartiniai DI generatyviniai modeliai sunkiai identifikuoja darbo naujumą, mokslinį pagrįstumą, inovatyvumą, originalumą ir pan.).

Studentų atsakomybės:

1. susipažinti su Vilniaus universiteto Akademinės etikos kodeksu, Universitete DI ir DI generatyvinių modelių naudojimą reguliuojančiais dokumentais ir DI generatyvinius modelius naudoti atsakingai ir etiškai;
2. jei leidžiama naudoti DI ir DI generatyvinius modelius studijų procese, atskleisti šio naudojimo faktą;
3. kilus klausimams dėl etiško naudojimo, konsultuotis su dalyko (modulio) dėstytoju ar rašto darbo vadovu;
4. laikytis Gairėse bei KAP metodiniuose nurodymuose bei dėstytojo pateiktuose reikalavimuose aprašytų DI citavimo ir DI panaudojimo deklaratavimo reikalavimų rašto darbuose.

Akademinė etika ir autorystė

DI turėtų būti naudojamas atsakingai, gerbiant asmeninius duomenis, nediskriminuojant ir skatinant teisingumą bei skaidrumą.

Universitete rašto darbą ar straipsnį vadovaudamasis teisės aktuose nustatytais reikalavimais turi rengti pats studentas ar akademinis darbuotojas, o šio principo nesilaikymas laikytinas galimu Vilniaus universiteto akademinės etikos kodekso pažeidimu, nes pažeidžia akademinio sąžiningumo principą, todėl toks pažeidimas gali būti skundžiamas akademinės etikos komisijoms teisės aktų nustatyta tvarka. DI generatyviniai modeliai ar bet kurie kiti DI modeliai, negali būti laikomi autoriais ar bendraautoriais, todėl DI neturėtų būti nurodomi kaip teksto (pvz., baigiamojo darbo, mokslinio straipsnio ar kitos akademinės publikacijos) ar kito kūrinio autoriai arba bendraautoriai. Kūrinio autoriumi gali būti tik žmogus.

DI generatyvinio modelio panaudojimas privalo būti atskleistas, todėl jei tekste ar kitame kūrinyje buvo naudotas DI generatyvinis modelis, tai būtina aiškiai nurodyti (atitinkamai cituojant ir (ar) pateikiant DI generatyvinio modelio naudojimo deklaraciją). DI generatyvinio modelio panaudojimo akademinėje darbe neatskleidimas laikytinas akademinio nesąžingumu.

Kuriant turinį DI generatyvinio modelio pagalba, gali kilti klausimų, kiek tokio modelio sukurtas rezultatas yra nulemtas konkrečių identifikuojamų žmonių indėlio. Šie atvejai turės būti sprendžiami individualiai, tačiau siūlytina atsižvelgti į tai, kad jeigu būtų nustatyta, jog žmonių indėlis neturi jokių tiesioginių sąsajų su gauto rezultato išraiška (pvz., apsiriboja tik bendru nurodymu sugeneruoti tam tikro pobūdžio kūrinių), jų laikyti gauto rezultato autoriais nebūtų galima. Kita vertus, autoriaus (žmogaus) kūrybiškumo išraiška gali būti ne tik sukurtas turinys, bet ir tokio turinio parinkimas ar išdėstymas / struktūravimas. Pastaruoju atveju, DI sugeneruoto turinio rinkinio autorystė galėtų priklausyti tokį parinkimą ir (ar) išdėstymą atlikusiam žmogui (žmonėms).

Citavimas

Naudojant DI sugeneruotą tekstą arba jo parafrazę, šis turinys turėtų būti identifikuojamas bei cituojamas panašiai, kaip ir kito autoriaus tekstas. Žemiau pateikti citavimo pavyzdžiai, tačiau svarbu atkreipti dėmesį, jog skirtingose disciplinose naudojamos skirtingas citavimo stilius, todėl reikėtų šiuos pavyzdžius atitinkamai adaptuoti.

1 citavimo pavyzdys. Pagal instrukciją „Kas yra *ChatGPT*?“, DI sugeneruotame tekste pateikiama, kad tai yra „vienas pažangiausių kalbos modelių, kuris gali suprasti ir generuoti natūralią kalbą, bandant atlikti konkrečias užduotis arba teikti bendravimo pagalbą vartotojams“ (OpenAI, 2023; žr. teksto transkripciją priede).

Jei DI sugeneruotas turinys yra perfrazuojamas, nuoroda į DI įrankį taip pat pateikiama taip, kaip ir naudojantis kito autoriaus mintimis:

2 citavimo pavyzdys. OpenAI (2023) pateikia atsakymą, kad *ChatGPT* yra vienas pažangiausių kalbos modelių.

Tais atvejais, kai darbe naudojama tik dalis sugeneruoto didesnės apimties DI generatyvinio modelio atsakymo į instrukciją, pilną sugeneruotą tekstą reikia pridėti darbo prieduose. Pilnas sugeneruotas tekstas svarbus dėl to, jog DI generatyviniai modeliai kiekvieną kartą generuoja unikalius atsakymus, todėl net pateikus tokią pačią užklausą iš naujo, sugeneruotas tekstas jau bus kitoks. Tokiu atveju tekste pakanka referuoti į priede pateiktą visą tekstą ir jį aprašyti apibendrintai, kaip nurodyta aukščiau pateiktuose pavyzdžiuose.

Taip pat DI įrankių sukurto turinio panaudojimas turi būti aprašomas ir literatūros sąrašė: nurodoma įrankį sukūrusi organizacija, teksto sukūrimo data, įrankio pavadinimas bei interneto svetainės adresas.

Nuorodos naudotų šaltinių sąrašė pavyzdys. Autorius. (Data). Įrankio pavadinimas (Įrankio versija) [Didysis kalbos modelis]. URL

OpenAI. (2023). *ChatGPT* (rugsėjo mėn. 6 d. užklausa) [Didysis kalbos modelis]. <https://chat.openai.com/chat>

Naudojimas mokslinėse publikacijose

Dėl DI generatyvinių modelių naudojimo mokslinėse publikacijose sprendžia patys mokslininkai laikydamiesi akademinės etikos normų ir atsižvelgdami į leidėjo ar finansuotojo nustatytas taisykles ar politiką. Kai kur tokių įrankių naudojimas yra draudžiamas, todėl svarbu laikytis leidėjo nustatytų taisyklių ar rekomendacijų.

Duomenų sauga

DI įrankių naudojimas siejamas su tokiais kibernetinėmis grėsmėmis kaip duomenų nutekinimas, asmens duomenų atskleidimas, neteisingos informacijos pateikimas ar kitos potencialios grėsmės. Siekiant nepažeisti duomenų saugumo į DI generatyvinius modelius negalima kelti konfidencialių duomenų ir (ar) informacijos, įskaitant, bet neapsiribojant asmens duomenimis, kitų originaliomis idėjomis, finansine, trečiųjų šalių ar kita jautria informacija, taip pat dar nepaskelbtais tyrimų duomenimis. DI generatyviniai modeliai negarantuoja šių duomenų saugumo, todėl jie gali būti atskleisti ir (ar) prieinami trečiosioms šalims, pažeisti asmens duomenis, asmenų intelektualinės nuosavybės teises ar galimybes siekti jų apsaugos (pvz., patentuoti).