Mokymasis su multimodalinio DI priemonėmis

Įvadas į pagrindines dirbtinio intelekto sąvokas

Autorė: Kacsur Annamária, ITStudy

Recenzentė: Hartyányi Mária, ITSTudy

M1-T1-MLC6

Mokymosi tikslas

Naujausi dirbtinio intelekto sprendimai gerokai pranoko tekstinio bendravimo galimybes. Didieji kalbos modeliai (ChatGPT, Gemini, Claude), iš pradžių sukurti tekstui generuoti, dabar gali ne tik generuoti vis kokybiškesnius vaizdus, vaizdo ir garso įrašus, bet ir apdoroti skaitmeninę informaciją iš bet kurio išorinio šaltinio, pavyzdžiui, jutiklių. Tokių išplėstų galimybių modeliai bendrai vadinami **multimodalinio dirbtinio intelekto** sistemomis.

Šioje mokomojoje medžiagoje pristatomos multimodalinio dirbtinio intelekto galimybės švietime ir paaiškinama, kuo jis pranoksta pastaraisiais dešimtmečiais plačiai paplitusią daugialypės terpės mokomąją medžiagą.

Kaip mokomės?

Mokslininkai jau seniai tyrinėja mokymosi prigimtį, tačiau vis dar tik iš dalies suprantame, kaip mokosi maži vaikai ir kaip įvairi patirtis, įgyta tyrinėjant aplinką, virsta nuosekliomis žiniomis. Kai mažas vaikas žaidžia su kamuoliu, regimieji įspūdžiai, lytėjimas, judesiai ir su jais susiję žodžiai susijungia į bendrą patirtį. Supratimas atsiranda ne iš paprastos šių įspūdžių sumos, o atpažįstant jų tarpusavio ryšius.



Kaip mokosi kūdikiai? (Šaltinis: ChatGPT)

Vyresni vaikai ir suaugusieji mokosi panašiai. Pasaulio nesuvokiame per atskirus „kanalus“: matome, girdime, veikiame, kalbame ir skaitome, o šią patirtį interpretuojame susiedami ją tarpusavyje ir su ankstesnėmis žiniomis. Tai atitinka konstruktyvistinę mokymosi teoriją, pagal kurią besimokantieji ne tik priima jiems perduodamą informaciją, bet ir interpretuoja ją remdamiesi savo patirtimi bei turimomis žiniomis, nustato ryšius ir patys kuria prasmę.

XX a. paskutiniame dešimtmetyje mokslininkai vis daugiau dėmesio skyrė tam, kaip skirtingi raiškos ir vaizdavimo būdai padeda suprasti informaciją. Taip susiformavo **multimodalumo**[1] teorija, o švietime pirmą kartą atsirado **multimodalinio mokymosi** ir **multimodalinės pedagogikos** sąvokos, pabrėžiančios, kad žinias taip pat galima kurti atpažįstant ir plėtojant skirtingų reprezentacijų ryšius.

Nuo daugialypės terpės prie multimodalumo

Kalba, ypač rašytinis tekstas ir sakytinė kalba, ilgą laiką buvo pagrindinė mokymo priemonė: mokytojas aiškina, o mokiniai skaito, užsirašo ir vėliau žodžiu arba raštu parodo, ką išmoko.



Tradicinis mokymas: mokytojas aiškina, mokiniai skaito, užsirašo ir vėliau atsiskaito žodžiu arba raštu (Šaltinis: ChatGPT)

Tobulėjant skaitmeninėms technologijoms, **daugialypė terpė** – bendras teksto, vaizdų, garso, vaizdo įrašų ir animacijos naudojimas – tapo vis prieinamesnė. Sąvokos „**būdas**“ arba „**modalumas**“ nusako, kokia forma pateikiama informacija arba kokio tipo duomenys patenka į sistemą. Tai gali būti tekstas, vaizdai, garsas, vaizdo įrašai ir išorinių jutiklių duomenys, pavyzdžiui, meteorologijos stotyje užfiksuota vidutinė paros temperatūra.

[1] Žodyje „multimodalumas“ priešdėlis „multi“ yra lotyniškos kilmės ir reiškia „daug“ arba „keli“; „modalumas“ kilęs iš lotyniško žodžio modus (būdas) ir šiame kontekste reiškia vaizdavimo, komunikacijos ar raiškos būdą.

Skaitmeninės sistemos gali apdoroti:

- tekstą;
- vaizdus ir grafiką;
- kalbą, muziką ir kitus garsus;
- vaizdo įrašus;
- judesius ir gestus aprašančius duomenis;
- jutiklių duomenis, pavyzdžiui, padėties, temperatūros ar greičio matavimus.

Tai nereiškia, kad modalumas sutampa su failo formatu. Pavyzdžiui, JPEG ir PNG yra skirtingi vaizdo failų formatai, tačiau abu priklauso tam pačiam **vaizdiniam modalumui**.



Daugialypė terpė mokyme: kelios medijos naudojamos kartu, o informacija pateikiama greta.

Pastarųjų dešimtmečių patirtis parodė, kad interaktyvūs mokomieji kompaktiniai diskai ir spalvingos pateiktys, papildytos vaizdais, animacija bei garsais, padeda mokytis, tačiau kokybinio supratimo pagerėjimo nesuteikia.

Daugumoje pamokų ir šiandien vyrauja tekstas bei sakytinė kalba (vadovėliai, mokytojo aiškinimas visai klasei ir sąsiuviniai).

	
Pateikimas keliais kanalais	**Prasmės kūrimas siejant skirtingas reprezentacijas**
DAUGIALYPĖ TERPĖ	MULTIMODALUMAS

Jei mokiniai ir toliau daugiausia priima jau parengtą informaciją, o skirtingomis medijomis perteikiamų prasmių susiejimas aiškinant visiškai priklauso nuo mokytojo iniciatyvos, medijų elementai perteikia tą pačią žinią, nepagilindami supratimo.

Taikant multimodalinį pedagoginį požiūrį, mokiniai ne tik gauna informaciją įvairiomis formomis, bet ir patys lygina, sieja, keičia bei kuria skirtingas reprezentacijas. Taigi dėmesys perkeliamas nuo informacijos perteikimo daugialype terpe prie mokinių multimodalinio prasmės kūrimo.



Multimodalinis žinių kūrimas (Šaltinis: ChatGPT)

Funkcijos formulė, jos grafikas ir realų judėjimą aprašantis pasakojimas nėra tiesiog trys skirtingi tos pačios mokomosios medžiagos „formatai“. Mokiniai turi atpažinti, kaip pasakojime aprašyti pokyčiai atsispindi grafike ir kaip juos galima išreikšti matematiniais sąryšiais. Prasmė kuriama atpažįstant reprezentacijų tarpusavio ryšius.

Multimodalinis dirbtinis intelektas

Multimodalinis dirbtinis intelektas kaip įvestį priima skirtingų tipų duomenis (tekstą, vaizdus, garsą ir vaizdo įrašus) ir pagal užklausą generuoja skirtingų tipų išvesties duomenis.



Multimodalinis dirbtinis intelektas (Šaltinis: ChatGPT)

Multimodalinis dirbtinis intelektas[2] viename modelyje gali apdoroti ir interpretuoti kelių tipų duomenis – tekstą, vaizdus, garsą, vaizdo įrašus ir kitus. Sąvoka „multimodalinis“ reiškia įvesties ir išvesties duomenų, kuriuos modelis gali apdoroti, įvairovę.

Jis ne tik sujungia skirtingus medijų formatus, bet ir „išmoksta“ jų tarpusavio ryšius

Taikymo sritys

Sveikatos priežiūra

Dirbtinis intelektas gali analizuoti rentgeno ar magnetinio rezonanso vaizdus kartu su rašytiniais paciento klinikiniais įrašais, bendrai interpretuodamas šiuos šaltinius nustatyti dėsningumus ir jais remdamasis pasiūlyti galimą diagnozę.



Multimodalinis DI sveikatos priežiūroje (Šaltinis: ChatGPT)

Klientų aptarnavimas ir prieinamumas

Multimodalinis DI leidžia regos ar judėjimo negalią turintiems naudotojams užduoti klausimus žodžiu ir prašyti atsakymų jiems tinkama forma.

Turinio kūrimas

[2] **Modalumo kodavimas**: kiekvieno tipo įvesties duomenis pirmiausia apdoroja specializuotas koduotuvas. Vaizdo koduotuvas vaizdo elementus (pikselius) paverčia skaitiniais duomenimis, kalbos koduotuvas tekstą suskaido į žetonus (tokenus), o garso koduotuvas iš kalbos ar garso išskiria svarbius požymius. Šie koduotuvai užfiksuoja esminius kiekvieno modalumo požymius ir atmeta nereikšmingas detales.

Multimodalinė integracija (suliejimas): modelis sujungia kelių modalumų informaciją, kad susidarytų vientisą įvesties duomenų vaizdą. Multimodaliniai pamatiniai modeliai šią integraciją paprastai įgyvendina naudodami bendrus transformerių sluoksnius ir dėmesio mechanizmus, todėl vieno modalumo informacija daro įtaką kito modalumo interpretavimui. Pavyzdžiui, kartu interpretuodamas žmogaus veido išraišką, balso toną ir ištartus žodžius, modelis gali numanyti jo emocijas ar ketinimus.

Tokie modeliai kaip ChatGPT ir Gemini, gavę bendrą teksto ir vaizdo įvestį, keičia vaizdą arba kuria naują kūrybinį turinį.

Robotika ir autonominės sistemos

Autonominės transporto priemonės ir robotinės sistemos naudoja multimodalinį dirbtinį intelektą kamerų vaizdams ir erdviniams duomenims sujungti, aplinkai suvokti ir navigacijos sprendimams realiomis sąlygomis priimti.

Dokumentų analizė

Gavęs dokumentą su diagramomis, schemomis ir įterptais vaizdais, multimodalinis dirbtinis intelektas apdoroja visus vaizdinius elementus kartu su tekstu ir atsako į klausimus, juos bendrai interpretuodamas.

Profesinio mokymo pavyzdžiai

Vykdydami aplinkosaugos projektą, mokiniai į DI sistemą įkelia upės būklę rodantį vaizdo įrašą ir dokumentą su matavimų rezultatais. DI padeda nustatyti ryšius tarp vaizdo įraše matomų reiškinių ir matavimo rezultatų bei išvadas, kurias galima padaryti kartu nagrinėjant abu šaltinius.



Multimodalinis DI aplinkosaugos projekte (Šaltinis: ChatGPT)

Per istorijos pamoką mokiniai gali prašyti DI atsakyti į jų klausimus kartu analizuojant to laikotarpio nuotrauką, plakatą ir rašytinį istorinį šaltinį.



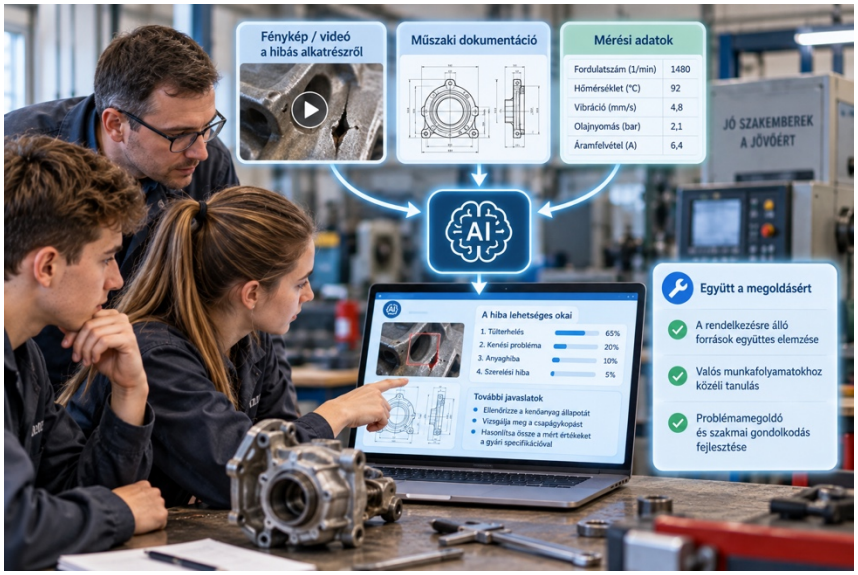
Multimodalinis DI istorijos pamokoje (Šaltinis: ChatGPT)

Per fizikos pamoką mokiniai gali kartu nagrinėti judėjimo vaizdo įrašą, atitinkamus matavimo duomenis ir grafiką: ar grafikas atitinka tai, ką matome įraše, ir kaip šį ryšį galima paaiškinti fizikos dėsniais?



Multimodalinis DI fizikos pamokoje (Šaltinis: ChatGPT)

Profesinio mokymo mokiniai įkelia sugedusios detalės nuotrauką ar vaizdo įrašą kartu su technine dokumentacija ir matavimo duomenimis. Bendrai analizuodamas šaltinius, DI gali atsakyti į mokinių klausimus apie galimas gedimo priežastis.



Multimodalinis DI pamokoje (Šaltinis: ChatGPT)

Gairės mokytojams

Šiuolaikinės dirbtinio intelekto programos leidžia mokytojams greitai ir lengvai kurti įspūdingus vaizdus, vaizdo įrašus, pateiktis ir infografikus. Rengti mokomąją medžiagą tampa lengviau, jos greitas senėjimas kelia mažiau sunkumų, o pritaikymas individualiems mokinių poreikiams gali tapti realybe.

Multimodalinis dirbtinis intelektas gali padėti pasiekti gilesnių pokyčių, jei skatinsime mokinius tarpusavyje lyginti pačių sukurtą ar DI sugeneruotą turinį, prašyti jį išanalizuoti, įvertinti jų pačių idėjas ir pagrįsti pateiktus atsakymus.

Multimodalinis DI atveria naujų aktyvaus mokymosi galimybių. Mokytojo padedami mokiniai gali tapti konstruktyviais savo mokymosi dalyviais. Mokytojo užduotis – suprojektuoti ir sukurti mokymosi aplinką, kurioje multimodalinis DI skatintų mokinius mąstyti, interpretuoti, lyginti ir patiems kurti prasmę. Šiandien šios technologijos gali sumažinti tekstinio informacijos apdorojimo dominavimą ir priartinti ugdymą prie tikrosios mokymosi prigimties.

Literatūros šaltiniai

- Simamkele Matuntuta (2026), How AI is Transforming Vocational Education and Training, straipsnis svetainėje cloudassess.com (žiūrėta: 3/7/2026
- Ayush Chauhan (2025): Multimodal AI: Beyond Text and Images, straipsnis svetainėje https://www.techaheadcorp.com,
- Kamalmeet Singh (2026): How Multimodal AI Works?, tinklaraščio straipsnis svetainėje https://medium.com, žiūrėta 8/7/2026
- Amol Yadhav (2024): Multimodal AI: The Next Frontier in Personalized Education, tinklaraščio straipsnis svetainėje https://medium.com/, žiūrėta 8/7/2026

- Peraschi Selvan Subramanian (2025): Multimodal AI: The future of integrated intelligence, straipsnis žurnale World Journal of Advanced Engineering Technology and Sciences, žiūrėta 9/7/2026

Nuorodos

Multimodalinis DI https://www.llamaindex.ai/glossary/multimodal-ai

Kaip veikia multimodalinio DI modeliai? YouTube vaizdo įrašas:

https://www.youtube.com/watch?v=WkoytlA3MoQ

Autorių teisės © 2026 VITA konsorciumas

Šia medžiaga galima dalintis ir ją adaptuoti pagal Creative Commons Priskyrimo — Analogiško platinimo 4.0 tarptautinės licencijos (CC BY-SA 4.0) sąlygas.

Finansuojama Europos Sąjungos. Išsakytos nuomonės priklauso tik autoriui(-iams) ir nebūtinai atspindi Europos Sąjungos ar Nacionalinės agentūros nuomonę.