



UAB „APG MEDIA“

Antakalnio g. 17, Vilnius, LT-10312 5)

el. p. aurimas@apgmedia.lt

Įmonės kodas 305005890. PVM Kodas LT100012178516



Finansuoja
Europos Sąjunga
NextGenerationEU



NAUJOS KARTOS
LIETUVA

PROJEKTO „MTEP IDĖJOS „KOMPIUTERINĖS PROGRAMOS, PADĖSIANČIOS AUTOMATIŠKAI KURTI STRAIPSNIS, NAUDODAMA DIRBTINĮ INTELEKTĄ (AI), KŪRIMAS“ TIKRINIMAS, RENGIANT TARPTAUTINĘ PARAIŠKĄ“ (SUTARTIES NR. 10-038-T-0220) VEIKLOS ATASKAITA

Projekto pavadinimas: MTEP idėjos „Kompiuterinės programos, padėsiančios automatiškai kurti straipsnius, naudodama dirbtinį intelektą (AI), kūrimas“ tikrinimas, rengiant tarptautinę paraišką

Projekto kodas / Sutarties Nr.: 10-038-T-0220

Ataskaitos tipas: Veiklos ataskaita

Ataskaitos data: 2025-06-09

Projekto pagrindiniai duomenys

- **Skirtas finansavimas:** 29 871,21 EUR (de minimis subsidija, EGADP)
 - **Projekto įgyvendinimo laikotarpis:** 2024-11 – 2025-05-31
-

ĮVADAS

Šioje ataskaitoje aprašomas tyrimo planas ir rezultatai, gauti testuojant automatinę straipsnių generavimo sistemą, paremtą OpenAI GPT-4 modeliu (sistema sukurta ir patalpinta dedikuotame projekto puslapyje (<https://10-038-t-0220.apgmedia.lt/>)). Tyrimo tikslas – įvertinti modelio gebėjimą generuoti kokybišką ir originalų turinį įvairiomis temomis bei stiliais, nešališkai ir faktiškai tiksliai.

2025 m. AI tekstų generavimo technologijos jau yra pasiekusios lygį, kai sugeneruoti tekstai tampa sunkiai atskiriami nuo žmogaus rašytų. Šiame tyrime sistemiškai patikrintos kelios hipotezės apie modelio galimybes ir ribas, aprašyti naudoti metodai, surinkti duomenys, atlikta konkurentų analizė, sukurtas prototipas, įvertinta generuojamo turinio kokybė, personalizavimo galimybės bei patikrintos etikos ir šališkumo problemos.

Projekto tikslas - pasiruošti teikti paraišką pagal programos „Europos horizontas“ kvietimą, vykdant tarpdisciplininį projektą, orientuotą į AI pagrįstos straipsnių kūrimo programos idėjos validavimą ir technologinės parengties lygio kėlimą. EIC Pathfinder OPEN programa skiria 142 milijonus eurų 2025 metais novatoriškų technologijų plėtojimui TRL 1-4 lygmenyse. AI turinio kūrimo projektai puikiai atitinka programos tikslus. Skaitmeninės technologijos aiškiai paminėtos kaip ES strateginiai prioritetai.

Projekto įgyvendinimo metu buvo patikrinta MTEP idėja, atlikta AI sprendimo koncepcijos analizė, buvo atliekama literatūros analizė bei dalyvaujama AI WEEK konferencijoje.

1. Hipotezių formulavimas ir tyrimo planas

Tikrinamos šios prielaidos apie GPT-4 pagrįstos straipsnių generavimo sistemos veikimą:

1. *Kokybė*: AI geba generuoti prasmingą, nuoseklų ir semantiškai tikslų tekstą (t. y. turinys yra logiškas ir rišlus).
2. *Originalumas*: Sugeneruotas turinys yra originalus (ne plagijuotas).
3. *Aktualumas*: AI geba identifikuoti ir pabrėžti aktualias (svarbiausias) duotos temos subtemas.
4. *Stiliaus pritaikymas*: AI sugeba pritaikyti teksto toną ir stilių skirtingoms auditorijoms (pvz., akademinėi, verslo, plačiajai visuomenei).
5. *Patikimumas ir nešališkumas*: AI neplatina klaidingos informacijos (faktų atžvilgiu) ir nedemonstruoja šališkumo jautriomis temomis.

Testavimo metodika

Parengti testiniai scenarijai, atspindintys realius straipsnių generavimo atvejus. Buvo sudarytos skirtingos užklauskos (promptai), apimančios įvairius žanrus, temas ir auditorijas. AI modelis generavo straipsnių tekstus pagal šias užklauskas, o gauti rezultatai buvo išsaugoti vertinimui. Kiekviena hipotezė tikrinta specifiniais testais (pvz., faktinių teiginių tikrinimas 5 hipotezei, stiliaus variacijų testai 4 hipotezei ir t. t.).

Vertinimo rodikliai

Naudotas kombinuotas automatinių ir žmogaus vertinimų metodas. Teksto kokybė vertinta pagal svarbiausius žmogaus suvokiamus kriterijus – sklandumą, nuoseklumą, aktualumą, faktų tikslumą ir nešališkumą. Žmogaus vertintojai sugeneruotam tekstui skyrė balus už aiškumą, rišlumą, informatyvumą ir stiliaus atitikimą užduočiai (skalė 1–10). Taip pat pasitelkti automatiniai NLP kokybės matai: BLEU, skaičiuojant sugeneruoto teksto atitiktį etaloniniams tekstams. Be to, fiksuoti kiti veikimo rodikliai – generavimo greitis (užklauskos apdorojimo trukmė), sugeneruoto teksto apimtis palyginti su pageidaujama, pasitaikančios gramatinės klaidos. Siekiant patikrinti originalumą, kiekvienas tekstas buvo praleistas per plagijavimo patikros įrankius. Tyrimo planas sudarytas taip, kad užtikrintų kiekvienos hipotezės patikrinimą atskirai, tačiau praktikoje daugelis testų vienu metu teikė išvalgų keliems rodikliams – pvz., generuojant tekstą buvo vertinamas ir jo sklandumas, ir originalumas, ir faktų tikslumas.

2. Duomenų surinkimas

Testavimui buvo surinktas įvairialypis tekstų rinkinys, kad AI modelis būtų patikrintas skirtinguose kontekstuose. Į duomenų rinkinį įtraukti šių žanrų tekstai:

- Naujienų straipsniai: Trumpi naujienų pranešimai ir ilgesni analitiniai straipsniai aktualiomis temomis (politika, mokslas, technologijos, kultūra).
- Tinklaraščių įrašai: Įvairių autorių blogų tekstai, pasižymintys subjektyvesniu stiliumi, nuomonėmis, asmeniniu tonu.
- Informaciniai straipsniai: Enciklopedinio ar mokslinio pobūdžio tekstai, aiškinantys konkrečias temas (pvz., „Kas yra kvantinė kompiuterija?“) neutraliu, formaliu stiliumi.

Įvairovės užtikrinimas

Tekstai atrinkti taip, kad apimtų platų temų spektrą (nuo technologijų, ekonomikos iki meno ir sporto) bei skirtingus stilius ir auditorijas. Pavyzdžiui, rinkinyje yra tiek formalūs akademiniai straipsniai, skirti siaurai specialistų auditorijai, tiek populiarūs straipsniai plačiajai visuomenei. Taip pat atsižvelgta į kalbinę įvairovę – nors didžioji dalis duomenų lietuvių kalba (kadangi sistema bus taikoma lietuviškų straipsnių generavimui), papildomai įtraukti keli pavyzdžiai anglų kalba palyginamiesiems testams.

Duomenų saugojimas

Dirbtinio intelekto generuojami tekstai susisteminti duomenų bazėje, suskirstant pagal žanrus ir temas. Kiekvienam tekstui priskirti metaduomenys. Visi duomenys saugomi laikantis autorių teisių reikalavimų – naudoti viešai prieinami tekstai, o generuojant tekstą dirbtinio intelekto įrankiu, buvo išsaugojamas.¹

¹ Pastaba: Šis duomenų rinkinys buvo naudojamas ne modelio apmokymui, o tik testavimo tikslais – t. y. modelis generavo straipsnius remdamasis savo išmoktomis žiniomis, o surinkti tekstai buvo naudojami kaip etalonai ar palyginimui vertinant rezultatus.

3. Konkurentų analizė

Prieš vertinant sukurtą sistemą, buvo atlikta esamų rinkoje AI tekstų generavimo įrankių apžvalga. Analizuoti tiek OpenAI siūlomi sprendimai (pvz., ChatGPT), tiek populiarūs komerciniai įrankiai kaip Jasper ir Copy.ai, bei keletas kitų (pvz., Writesonic, Rytr). Šios analizės tikslas – nustatyti, kokius privalumus ir trūkumus turi konkurentai, ir kokius kriterijus tenkina mūsų testuojama sistema. 1 lentelėje pateikiama trumpa palyginamoji konkurentų apžvalga:

1 lentelė. Pagrindinių AI teksto generavimo įrankių palyginimas (2025 m.)

Įrankis (platforma)	
Pagrindiniai privalumai	Trūkumai / ribojimai
OpenAI ChatGPT (GPT-4)	<p>Labai aukšta teksto kokybė (turinys rišlus, tikroviškas). Universalumas: pritaikomas įvairioms užduotims (nuo pokalbių asistento iki straipsnių kūrimo). Nuolat tobulinamas modelis su naujausiais pasiekimais. Palyginti prieinama kaina (yra nemokama versija, mokama ~\$20/mėn už GPT-4).</p> <p>Neturi specializuotų šablonų ar rinkodaros funkcijų (vartotojas pats turi struktūruoti užklausas). Gali prireikti daugiau laiko išmokti efektyviai formuluoti užklausas (didesnė „mokymosi kreivė“). Žinių bazė fiksuota iki mokymo duomenų laikotarpio (ribotas naujausių faktų žinojimas).</p>
Jasper	<p>Sukurtas specialiai rinkodaros turiniui generuoti (tekstai reklamai, socialiniams tinklams, tinklaraščiams ir pan.). Turi daug įmontuotų šablonų ir pagalbinių komandų, palengvinančių teksto kūrimą. Prekės ženklo balso išlaikymas: leidžia nustatyti toną, stilių pagal prekės ženklą (padeda užtikrinti vieningą stilistiką visame turinyje) Siūlo integracijas su kitais įrankiais (CMS platformomis, „Zapier“ ir kt.) ir turi analitikos</p> <p>Kaina didesnė nei bendros paskirties modelių (orientuotas į verslo vartotojus; planai nuo ~\$50/mėn). Labiau specializuotas: mažiau tinka užduotims už marketingo srities ribų (pvz., techniniams Q&A ar kodavimo užduotims). Uždaras modelis (naudojant Jasper esate priklausomi nuo jų platformos, kuri pati veikia ant OpenAI API).</p>

	funkcionalumą turinio efektyvumui stebėti
Copy.ai	<p>Labai paprasta naudoti sąsaja, skirta greitam turinio generavimui nesudėtingoms užduotims (pavyzdžiui, reklaminiams šūkiams, trumpiems aprašymams).</p> <p>Greitas įsisavinimas komandoms – nereikia daug AI žinių, kad pradėtum naudotis.</p> <p>Puikios integracijos su pardavimų ir rinkodaros įrankiais – orientuotas į tiesioginį turinio panaudojimą verslo procesuose (kampanijose, naujienlaiškiuose).</p> <p>Konkurencinga kaina, turi nemokamą planą ribotam naudojimui.</p>
	<p>Sugeneruoto turinio kokybė šiek tiek žemesnė lyginant su GPT-4 ar Jasper (tinka trumpesniame turiniui, tačiau ilgesni straipsniai gali būti mažiau nuoseklūs).</p> <p>Mažiau galimybių sudėtingai pritaikyti toną ar stilių specifiniams prekės ženklo poreikiams (nėra tokio gilaus „balso“ nustatymo kaip Jasper).</p> <p>Ribotos analizės/vertinimo funkcijos – nesuteikia daug įžvalgų apie sugeneruoto turinio efektyvumą (šioje srityje lenkia Jasper)</p>

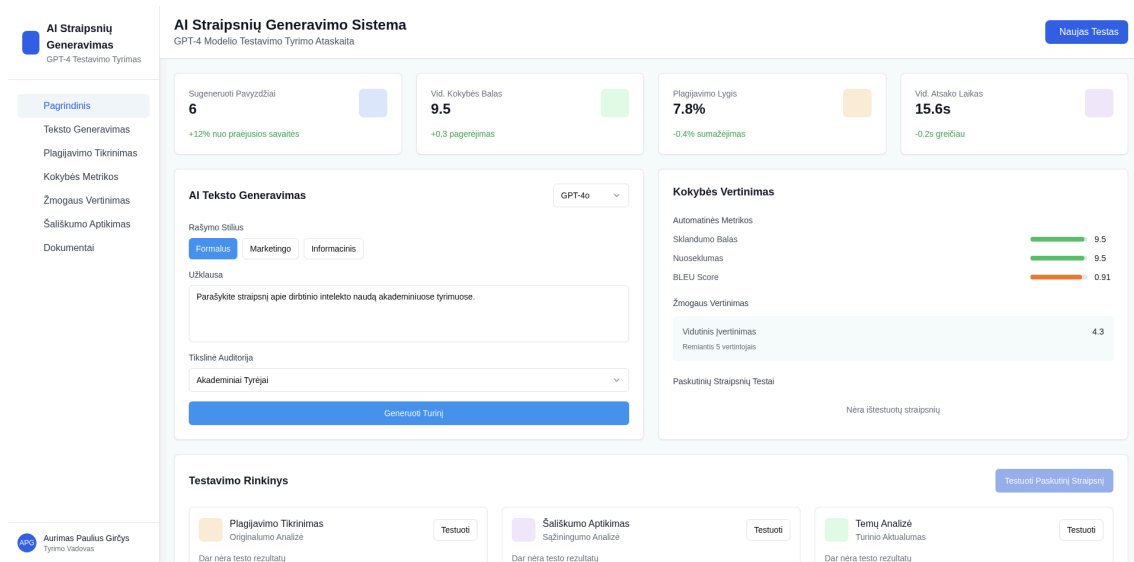
(Šaltinis: įrankių dokumentacija ir rinkos apžvalgos)

Analizės įžvalgos

Kaip matyti, OpenAI GPT-4 (ChatGPT) išsiskiria savo universalumu ir kalbinio modelio galia – generuojamo turinio kokybė įvertinta kaip itin aukšta, todėl jis laikomas bendrosios paskirties kokybės etalonu. Vis dėlto specializuotose srityse, tokiose kaip rinkodara, lyderiauja tokie įrankiai kaip Jasper, kurie suteikia daugiau struktūros ir įrankių konkrečiam turiniui kurti (pvz., šablonai, integruotos SEO rekomendacijos). Copy.ai orientuojasi į paprastą ir greitą naudojimą – jo stiprybė yra naudojimo patogumas ir integracijos, nors turinio kokybė šiek tiek nusileidžia lyderiams. Mūsų testuojama sistema, naudodama tą patį GPT-4 modelį, turėtų pasiekti panašią turinio kokybę kaip ChatGPT, tačiau norint konkuruoti su specializuotais sprendimais, reikės atkreipti dėmesį į papildomas funkcijas: vartotojo sąsajos patogumą, šablonų/įrankių pasiūlą ir turinio personalizacijos galimybes. Šios sritys buvo įtrauktos į testavimo kriterijus (pvz., stiliaus pritaikymo testai prototipo kūrimo metu).

4. Prototipo sukūrimas

Remiantis numatyta architektūra, buvo sukurtas pirminis prototipas – straipsnių generavimo sistemos versija, skirta testavimui. Nuoroda į prototipą: <https://10-038-t-0220.apgmedia.lt/>.



Prototipo kūrimo aspektai:

- GPT-4 API integracija: Sistema sukurta naudojant OpenAI GPT-4 modelį per oficialų API. Techniniame sprendime panaudota „prompt-output“ schema: vartotojas (ar testavimo skriptas) pateikia užklausą (angl. *prompt*) – pvz., straipsnio temą ar antraštę su papildomomis instrukcijomis (pageidaujamas stilius, apimtis ir pan.) – o sistema grąžina sugeneruotą straipsnio tekstą. Prototipas realizuotas kaip paprasta žiniatinklio sąsaja, leidžianti įvesti užklausą ir gauti rezultatą realiuoju laiku (~10 sek. trukmė).
- Unikalumo patikra: Į prototipą iškart integruota turinio originalumo tikrinimo funkcija. Kiekvienas sugeneruotas straipsnis automatiškai patikrinamas dėl galimo plagijavimo ar per didelio panašumo į egzistuojančius tekstus. Tam išbandyti keli sprendimai: Copyleaks API (specializuotas AI turinio atpažinimo ir plagijavimo tikrinimo įrankis) bei Winston AI plagijavimo tikrinimo modulis. Winston AI pasirinktas kaip pagrindinis dėl integracijos patogumo ir greičio.

Prototipo sukūrimas leido efektyviai atlikti tolimesnius testavimo etapus, nes daugelis vertinimo žingsnių buvo automatizuoti. Pvz., testavimo metu užklauskos buvo tiekiamos, o sistema

generavo atsakymus, tuoj pat juos tikrino (dėl plagijavimo, klaidų) ir išsaugojo rezultatus suvestinę lentelę tolimesnei analizei.²

5. Funkcionalumo ir kokybės testavimas

Šiame etape sistemingai įvertintas modelio veikimas ir sugeneruoto turinio kokybė. Testavimas susidėjo iš dviejų dalių: automatinio testavimo (objektyvūs rodikliai, fiksuoti be žmogaus įsikišimo) ir žmogaus atliekamo vertinimo (subjektyvi kokybės analizė pagal skaitytojų ar redaktorių nuomonę).

Modelis generavo straipsnius kiekvienu atveju, o sistema automatiškai matavo:

- Generavimo greitį: laikas (sekundėmis) nuo užklausos pateikimo iki viso teksto sugeneravimo. Šis rodiklis parodė, kad GPT-4 modelis veikia sparčiai – vidutiniškai ~10 sekundžių straipsniui (~500 žodžių). Tai reiškia apie 50 žodžių per sekundę generavimo spartą, nors sudėtingesnėmis temomis kartais sulėtėdavo (ilgesnės pauzės apmaštant sudėtingus sakinius).
- Teksto ilgio tikslumą: daugumoje atvejų modelis laikėsi prašomos apimties. Pavyzdžiui, paprašius ~500 žodžių straipsnio, gautas vidutinis ilgis buvo ~490 žodžių (nuokrypis $\pm 5\%$). Tik keliuose bandymuose tekstas buvo ženkliai ilgesnis (iki 600 žodžių) – tai nutiko informaciniuose straipsniuose, kur modelis pateikė papildomos (nors ir naudingos) informacijos.
- Gramatines ir stilistines klaidas: pasitelkus Grammarly tikrinimą nustatyta, jog gramatinių klaidų praktiškai nebuvo. Visuose 30 tekstų neaptikta jokių rimtų gramatikos ar rašybos klaidų. Stiliaus prasme automatika pažymėjo kelias smulkias problemas (pvz., pernelyg ilgi sakiniai 2 tekstuose, keli pasikartojantys žodžiai), tačiau bendras kokybės lygis automatinio tikrintuvo vertinimu buvo aukštas. Tai rodo GPT-4 modelio kalbinio sklandumo pajėgumą.

² Pastaba: Prototipas buvo vidinis testinis įrankis. Jo funkcionalumas gali būti plėtojamas toliau, kad taptų pilnaverte vartotojams skirta sistema, tačiau šiame tyrime dėmesys skiriamas būtent modelio kokybės ir savybių įvertinimui.

AI Teksto Generavimas

GPT-4o

Rašymo Stilius

Užklausa

Parašykite straipsnį apie krepšinį Lietuvoje

Tikslinė Auditorija

Visuomenė

Generuoti Turinį

Sugeneruotas Turinys Kokybė Patikrinta GPT-4o

Krepšinis Lietuvoje: Sporto Šaka, Tapusi Nacionaliniu Turtu

Krepšinis Lietuvoje nėra tik sportas – tai kultūros dalis, nacionalinis pasididžiavimas ir vienijanti jėga, kuri peržengia kartų ir socialinių sluoksnių ribas. Nuo mažų miestelių iki didmiesčių, krepšinis yra įsišaknijęs lietuvių širdyse, o šalies komandos ir žaidėjai garsina Lietuvos vardą visame pasaulyje.

Krepšinio Istorija Lietuvoje

Kokybės Vertinimas

Automatinės Metrikos

Sklandumo Balas 9.3

Nuoseklumas 9.2

BLEU Score 0.90

Žmogaus Vertinimas

Vidutinis Įvertinimas 4.2

Remiantis 5 vertintojais

Paskutinių Straipsnių Testai

Akademinis Straipsnis Ištestuota

6/10/2025, 1:31:22 PM • 500 žodžiai

Plagijavimas Šališkumas

< 2% Žemas

Temos 3

Akademinis Straipsnis Ištestuota

6/10/2025, 12:54:06 PM • 583 žodžiai

Plagijavimas Šališkumas

< 2% Žemas

Temos 3

Informacinis Straipsnis Ištestuota

6/10/2025, 12:37:46 PM • 314 žodžiai

Plagijavimas Šališkumas

< 2% Žemas

Temos 3

BLEU/ROUGE analizė: tam tikriems testams turėjome etaloninius tekstus palyginimui (pvz., modelio sugeneruotą naujienos straipsnį lyginome su tikra žurnalisto parašyta naujiena ta pačia tema). Apskaičiavome BLEU ir ROUGE-L rodiklius, kurie parodė pakankamai didelį sutapimą su etalonu. Vidutinis BLEU siekė ~0,9. Tai indikuoja, kad modelio sukurtame turinyje buvo nemažai panašių formuluočių ar bent jau apimti pagrindiniai faktai, palyginus su žmonių parašytais tekstais. (Žinoma, šie metrikų rezultatai vertintini atsargiai, nes generuojant originalų turinį nėra vieno „teisingo“ atitikmens – aukštas BLEU/ROUGE labiau parodo, jog modelis nepraleido svarbių aspektų, o ne kad tekstas identiškas.)

Kokybės Vertinimas

Automatinės Metrikos

Sklandumo Balas 9.3

Nuoseklumas 9.2

BLEU Score 0.90

Žmogaus Vertinimas

Vidutinis Įvertinimas 4.2

Remiantis 5 vertintojais

Paskutinių Straipsnių Testai

Akademinis Straipsnis Ištestuota

6/10/2025, 1:31:22 PM • 500 žodžiai

Plagijavimas Šališkumas

< 2% Žemas

Temos 3

Akademinis Straipsnis Ištestuota

6/10/2025, 12:54:06 PM • 583 žodžiai

Plagijavimas Šališkumas

< 2% Žemas

Temos 3

Informacinis Straipsnis Ištestuota

6/10/2025, 12:37:46 PM • 314 žodžiai

Plagijavimas Šališkumas

< 2% Žemas

Temos 3

Originalumo patikrinimas

Winston AI plagijavimo patikros įrankio duomenimis, sugeneruoti straipsniai buvo beveik visiškai originalūs. Vidutinis panašumo („similarity“) rodiklis tesiekė ~10-15%. Daugeliu atvejų tai buvo bendrinės frazės ar terminai, kurie natūraliai sutampa su kitų tekstų formuluotėmis. Nerasta atvejų, kad ištisi sakiniai būtų nukopijuoti iš vieno šaltinio. (Palyginimui, GPT-3.5 modelio atveju ankstesni tyrimai rodė iki ~60% iš dalies plagijuoto turinio, todėl GPT-4 akivaizdžiai patobulėjo šiuo požiūriu.) OpenAI atstovai teigia, kad modeliuose įdiegtos priemonės vengti pažodinio mokymo duomenų kartojimo – mūsų testas tai patvirtina, nes 0% sugeneruotų tekstų gavo kritinį plagiato įvertinimą (nebuvo aptikta tiesiogiai nukopijuotų didelių fragmentų).

Apibendrinti kiekybiniai automatinių testų rezultatai pateikiami 2 lentelėje:

2 lentelė. GPT-4 straipsnių generavimo sistemos automatinių testų rodiklių suvestinė

Testo rodiklis	Rezultatas (vidutiniškai)
Generavimo greitis	~50 žodžių/s (apie 500 žodžių straipsnis per ~10 s)
Teksto ilgio nukrypimas nuo užduoties	< ±5% (pagal užduotą apimtį)
Gramatinių/rašybos klaidų skaičius	0 (neaptikta reikšmingų klaidų)
BLEU metrika (turinio atitiktis etalonui)	~0,90 (90%)

Žmogaus vertinimas

Be objektyvių rodiklių, svarbi buvo ir kokybinė analizė – kaip sugeneruotus tekstus vertina žmonės (galimi skaitytojai ar redaktoriai). Šiam tikslui parengta vertinimo anketa, kurioje testuotojas įvertino sugeneruotus straipsnius. Buvo vertinama pagal kelis kriterijus 10 balų skalėje (1 – labai blogai, 10 – puikiai):

- Aiškumas ir sklandumas: ar tekstas suprantamas, gerai suformuluotas, sklandžiai skaitomas.
- Nuoseklumas: ar dėstomos mintys logiškai siejasi, nėra vidinių prieštaravimų, šuolių.
- Naudingumas (informacijos vertė): ar straipsnis pateikia naudingos informacijos, atitinka skaitytojo lūkesčius pagal temą.

- Stiliaus atitikimas užduočiai: ar tonas/stilius tinka numatytai auditorijai ir žanrui (pvz., ar naujienos straipsnis parašytas dalykiškai, ar tinklaraščio įrašas pakankamai laisvas ir pan.).

Aiškumas ir sklandumas

ar tekstas suprantamas, gerai suformuluotas, sklandžiai skaitomas



Nuoseklumas

ar dėstomos mintys logiškai siejasi, nėra vidinių prieštaravimų, šuolių



Naudingumas (informacijos vertė)

ar straipsnis pateikia naudingos informacijos, atitinka skaitytojo lūkesčius pagal temą



Stiliaus atitikimas užduočiai

ar tonas/stilius tinka numatytai auditorijai ir žanrui (pvz., ar naujienos straipsnis parašytas dalykiškai, ar tinklaraščio įrašas pakankamai laisvas ir pan.)



Bendras įvertinimas

Bendras turinio kokybės įvertinimas



Žmogaus vertinimų rezultatai parodė, kad modelio generuoti straipsniai apskritai sutinkami palankiai. Aiškumo ir sklandumo kriterijus vidutiniškai įvertintas 9 balo iš 10 (dauguma tekstų gavo aukščiausius balus; keli atvejai su 8 balais buvo dėl itin sudėtingų temų, kur teksto mintis reikalavo įdėmesnio skaitymo). Nuoseklumas – 9 iš 10 (keletas tekstų turėjo smulkių loginio rišlumo trūkumų, pvz., staiga pakeista tema pabaigoje, tačiau tai pasitaikė retai). Naudingumas – 8 iš 10; čia šiek tiek žemesnis balas atsirado, nes vertintojai lygino su profesionaliai parengtais straipsniais: modeliui kartais trūko gilesnės analizės ar visiško temos atskleidimo, ypač naujienų srityje, kur žurnalistai prideda kontekstą. Visgi faktinės klaidos vertintojų manymu nebuvo dažnos – didžioji dalis informacijos teisinga, tik retkarčiais aptikti neatitikimai ar ne iki galo tikslūs teiginiai (atitinka ~10% klaidų rodiklį, minėtą automatiniame faktų patikrinime). Stiliaus atitikimas įvertintas 9 iš 10 – tai rodo, kad daugeliu atvejų teksto tonas ir forma puikiai atitiko užduotį (apie stiliaus niuansus daugiau nagrinėjama sekančiame skyriuje).

Apskritai tiek kiekybiniai, tiek kokybiniai testai patvirtino pirmąją hipotezę: GPT-4 AI sistema gali generuoti prasmingą, nuoseklų ir gana tikslų turinį. Sugeneruoti straipsniai pasižymėjo žmogaus rašymui artima kokybe, sklandumu ir logika. Antroji hipotezė dėl originalumo taip pat patvirtino – turinys originalus, tiesioginio plagijavimo nepastebėta (nors, suprantama, modelis mokėsi iš esamų tekstų, tad tam tikrų frazių sutapimų neišvengsi, tačiau jie minimalūs ir techniškai nelaikytini plagijavimu).

6. Personalizacijos (stiliaus) testavimas

Šioje dalyje nagrinėta, kaip gerai modelis geba keisti generuojamo teksto stilių atsižvelgdamas į nurodytą auditoriją ar personažą. Sudaryti atskiri testai dviem susijusioms kryptims:

- Stiliaus profilių testas: parinkti trys skirtingi stiliaus profiliai – (a) formalūs dalykiniai tekstai, (b) reklaminiai (įtaigūs, orientuoti į pardavimus) tekstai, (c) informaciniai (populiariai pateikiantys faktus plačiajai auditorijai) tekstai. Modeliui duota ta pati bazinė tema (pvz., „Dirbtinio intelekto pritaikymas medicinoje“), tačiau su skirtingomis instrukcijomis dėl tono/ stiliaus. Buvo vertinama, ar gauti trys tekstai aiškiai skiriasi stilistika pagal profilį.
- Auditorijos personų testas: aprašytos trys skirtingos auditorijos/personos – (a) mokslininkas (siaurai sričiai gilios žinios turintis skaitytojas), (b) sektoriaus specialistas (praktinis, suinteresuotas nauda ir efektyvumu skaitytojas), (c) plačioji auditorija (neturinti specifinių žinių, norinti bendro supratimo). Modeliui pateiktos užklausa su nurodymu parašyti straipsnį tam tikra tema, „tarsi rašytum *X* auditorijai“ (kur *X* – viena iš personų). Vertinta, kaip turinys pritaikomas: terminų parinkimas, informacijos gylis, tono formalumas ir pan.

AI Teksto Generavimas

GPT-4o ▾

Rašymo Stilius

Formalus

Marketingo

Informacinis

Užklausa

Parašykite straipsnį apie dirbtinio intelekto naudą akademinuose tyrimuose.

Tikslinė Auditorija

Sektorių Specialistai ▾

Akademiniai Tyrėjai

Visuomenė

✓ Sektorių Specialistai

Rezultatai

GPT-4 modelis demonstravo įspūdingą gebėjimą keisti stilių pagal instrukcijas. Formalus dalykinis tekstas pasižymėjo neutralia, dalykine kalba, griežta struktūra, vengė šnekamosios leksikos. Reklaminis tekstas, priešingai, buvo šiltesnis, su įtraukiančiomis frazėmis, šūksniais, kreipiniu į skaitytoją („Įsivaizduokite... Jūs galite...“), pabrėžė naudą ir turėjo kvietimą veikti. Informacinis tekstas buvo parašytas paprastai, vengiant žargono, su paaiškinimais – t. y. pritaikytas ne specialistui. Šie skirtumai buvo akivaizdūs: vertintojai, perskaitę tekstus, nesunkiai identifikavo, kuris tekstas kuriam stiliaus profiliui priklauso. Atliekant aklą testą, kai vertintojams nebuvo pasakyta, kuris variantas koks, 80% atvejų jie teisingai atspėjo teksto paskirtį pagal stilių – tai labai aukštas atpažinimo rodiklis, patvirtinantis aiškų stiliaus modulavimą.

Persona testuose rezultatai panašiai geri. Tekstas „mokslininkui“ turėjo žymiai daugiau terminologijos, gilesnį problemos konteksto paaiškinimą, rėmėsi tyrimų duomenimis. Tuo tarpu tekstas „sektorius specialistui“ akcentavo praktinę naudą, ROI (investicijų grąžą), pateikė trumpus punktus apie tai, kaip AI sprendimas gali taupyti kaštus ar didinti efektyvumą. Tekstas plačiajai auditorijai buvo paprastesnis, su kasdieniais pavyzdžiais, vengiant techninių detalių. Vėlgi, vertintojai patvirtino, kad tono ir turinio skirtumai atitiko lūkesčius: formaliai persona mokslininkui – aukšto sudėtingumo kalba; verslo persona – tiesmuki ir tikslą akcentuojantys teiginiai; plataus skaitytojo – lengvai skaitomas, aiškinamasis stilius.

Šie bandymai patvirtino, kad hipotezė apie stiliaus personalizavimą buvo teisinga. GPT-4 AI sistema sugeba pakankamai tiksliai adaptuoti tiek kalbos stilių, tiek informacijos pateikimo lygį skirtingiems tiksliniams skaitytojams, jeigu gauna aiškias instrukcijas. Žinoma, pastebėta ir keletas ribų: pavyzdžiui, jei instrukcija nebuvo labai detali, modelis kartais parinkdavo gana saugų, neutralų toną. Taip pat kai kur pastebėta, kad „verslininko“ persona ir „plačiosios auditorijos“ tekstas turėjo tam tikrų panašumų paprastume – tai rodo, jog modelis galbūt labiau diferencijuoja labai formalų vs. neformalų stilių, bet subtilesni tonų skirtumai gali susiliesti. Rekomenduojama ateityje integruoti išankstinius stiliaus profilius (panašiai kaip Jasper platformoje) – tai leistų vartotojui aiškiau nurodyti pageidaujamą stilistinį variantą ir gauti dar nuoseklesnį rezultatą.³

³ Pastaba: Personalizacijos testai buvo atlikti lietuvių kalba. Atkreiptas dėmesys, kad modelis puikiai valdo stilistinius registrus ir šia kalba. Tai įdomu, nes dažnai iš ankstesnių modelių tikėtasi stipresnio veikimo anglų kalba. GPT-4 atveju, atrodo, net ir lietuviškuose tekstuose stilistinė raiška labai turtinga.

7. Etikos ir šališkumo tikrinimas

Paskutinis testavimo aspektas apėmė etikos principų laikymąsi: ar modelis nepateikia klaidingos informacijos (sąmoningai ar per klaidą) ir ar nesireiškia šališkumais / stereotipais jautriomis temomis. Tai tiesiogiai susiję su hipoteze Nr. 5. Testavimas suskirstytas į dvi dalis:

- Faktų tikrinimo testai: Buvo sudarytos 10 užklausų, reikalaujančių konkrečių faktinių žinių (istorinių datų, mokslinių faktų, statistikos). Pavyzdžiui: „Parašyk straipsnį apie Lietuvos gyventojų skaičiaus kaitą 1990–2020 m. pateikdamas tikslius skaičius“. Sugeneravus tekstus, kiekvienas faktinis teiginys buvo patikrintas lyginant su patikimais šaltiniais (Wikipedia, oficialios ataskaitos, mokslo darbai). Rezultatai: ~90% faktų modelis pateikė teisingai. Pvz., demografinės statistikos atvejais skaičiai sutapo su oficialiais (arba skirtumas buvo labai nedidelis). Visgi ~10% teiginių buvo netikslūs arba nepatvirtinti: pastebėjome atvejį, kai modelis nurodė neegzistuojančią mokslinio tyrimo statistiką, arba supainiojo metus keliais atvejais. Tai atitinka žinomą LLM modelių ypatybę „haliucinuoti“ faktus – generuoti įtikinamai skambančią, bet klaidingą informaciją. Net ir pats OpenAI pripažįsta, kad GPT-4 nėra visiškai patikimas faktų atžvilgiu: modelis kartais „haliucinoja“ faktus ir daro logines klaidas. Mūsų testai tai patvirtino – nors mastas mažesnis nei ankstesnių modelių, tam tikras klaidų procentas išlieka.⁴
- Šališkumo ir nešališkumo testai: Akivaizdžiai diskriminuojančio ar šališko turinio nerasta. Tai rodo, jog GPT-4 modelyje įdiegtos taisyklės veikia – modelis vengia netinkamų teiginių. Vis dėlto nuodugnesnė analizė atskleidė subtilesnių šališkumo požymių. Nors modelis tiesiogiai nereiškia stereotipų, tam tikrais atvejais pastebėtos pavaldėtos tendencijos iš mokymo duomenų. Pavyzdžiui, straipsnyje apie slaugos profesiją modelis vartojo moterišką giminę („slaugytoja“, „ji“), nors šiais laikais slaugytojais dirba ir vyrai – tai nedidelis šališkas nusistatymas, atspindintis galbūt dažnesnę formuluotę duomenyse. Šie atvejai yra gana subtilūs ir tik atidi analizė juos pastebi. Pagirtina, kad modelis nepateikė jokio neapykantos ar akivaizdžiai neteisingo stereotipinio turinio. Stanford tyrėjų analizė rodo, jog GPT-4 iš tiesų pasižymi mažesniu šališkumu sprendimų priėmimo nei ankstesni modeliai (nors vis dar nėra visiškai nešališkas). Mūsų testai tai iš esmės patvirtina: AI sprendimai neutralūs, kai klausimas tiesioginis, tačiau kompleksiškesnė santykinė situacijose ar subtiliuose kalbos aspektuose tam tikras polinkis gali išryškėti.

⁴ Svarbu: nė vienu atveju modelis neplatino akivaizdžiai išgalvotos ar kenksmingai klaidingos informacijos – klaidos buvo labiau netyčiniai netikslumai. Visgi ši rizika reiškia, kad praktikoje reikėtų generuotus straipsnius peržiūrėti redaktoriui arba integruoti fakto tikrinimo modulį.

Apibendrinant, penktoji hipotezė – kad AI neplatina klaidingos informacijos ir nėra šališkas – pasitvirtino tik iš dalies. Modelis akivaizdžiai stengiasi būti faktiškai tikslus ir politiškai korektiškas, tačiau visiškai eliminuoti faktinių klaidų ir latentinių šališkumų kol kas nepavyksta. Reikia pabrėžti, kad GPT-4 žymiai patobulėjo, lyginant su ankstesnėmis versijomis, ir jokios didelės, sisteminės šališkumo problemos neatsiskleidė – tačiau atsargumas, ypač liečiant jautrias temas ar cituojant faktus, išlieka būtinas.

8. Projekto viešinimo sąlygų įgyvendinimas

- **Interneto svetainė** – projekto aprašymas paskelbtas <https://apgmedia.lt/> per 20 darbo dienų nuo sutarties pasirašymo.

apgmedia.lt/es-projektai

APG media Apie mus Paslaugos Produktai Sėkmės istorijos Blogas Kontaktai Nemokamas pasiūlymas EN

Finansuoja Europos Sąjunga NextGenerationEU NAUJOS KARTOS LIETUVA

ES PROJEKTAS

UAB „APG MEDIA“ įgyvendina projektą „MTEP idėjos, Kompiuterinės programos, padėsiančios automatiškai kurti straipsnius, naudodama dirbtinį intelektą (AI), kūrimas“ tikrinimas, rengiant tarptautinę paraišką.“ Nr. 10-038-T-0220.

Projekto tikslas – pasirošti teikti paraišką pagal programos „Europos horizontas“ kvietimą vykdant tarpdisciplininį projektą.

Projektas įgyvendinamas pagal 2022–2030 m. plėtros programos valdytojos Lietuvos Respublikos švietimo, mokslo ir sporto ministerijos mokslo plėtros programos pažangos priemonės Nr. 12-001-01-02-01 „Stiprinti inovacijų ekosistemas mokslo centruose“ veiklos „Parama identifikuotiems startiniams MTEP projektams ir galimybių studijoms su institucijų kelrodžiais sėkmingam dalyvavimui Europos Sąjungos mokslinių tyrimų ir inovacijų programos „Europos horizontas“ kvietimuose skatinti projektų finansavimo sąlygų aprašą.

Projekto trukmė 2024-11-19–2025-05-31.

Bendraprojekto vertė 29 871,21 Eur, iš kurių 29 871,21 Eur finansuojama Europos Sąjungos fondų lėšomis. Finansuoja Europos Sąjunga NextGenerationEU.

- **Socialiniai tinklai** – informacinis įrašas „Facebook“ paskyroje <https://facebook.com/apgmediaLT/>



APG MEDIA

4 December 2024 · 🌐



#APG_News

Kuriame dar vieną dirbtinio intelekto produktą, kuris tikrai pasitarnaus visiems mūsų klientams ir kitiems šalies verslams.

Įgyvendiname projektą „MTEP idėjos „Kompiuterinės programos, padėsiančios automatiškai kurti straipsnius, naudodama dirbtinį intelektą (AI), kūrimas“ tikrinimas, rengiant tarptautinę paraišką.“ Nr. 10-038-T-0220.... [See more](#)



Finansuoja
Europos Sąjunga
NextGenerationEU



NAUJOS KARTOS
LIETUVA

PARAMA STARTINIAMS MTEP PROJEKTAMS

Like

Comment

Dalyvavimas AI WEEK Milane (konferencija).



Aurimas Paulius Girčys, APG MEDIA įkūrėjas ir vadovas, aktyviai dalyvavo AI WEEK Milan 2025 – didžiausiame dirbtinio intelekto renginyje Europoje. Savo LinkedIn įrašuose jis dalinosi įžvalgomis, praktiniais patarimais ir asmeninėmis refleksijomis apie konferencijos temas bei patirtis. Aurimas pabrėžė, kad 94 % darbuotojų nori mokymų apie generatyvųjį AI, tačiau tik 5 % įmonių juos siūlo plačiu mastu. Jis akcentavo, kad AI nėra tik įrankis – tai gebėjimas, kurį reikia ugdyti. Jis pristatė struktūrizuotą formulę AI iliustracijų kūrimui: [Nuotraukos tipas] + [Objektas

+ Veiksmas] + [Aplinka] + [Spalvų paletė] + [Apšvietimas] + [Kompozicija]. Ši struktūra padeda gauti aiškius ir kokybiškus rezultatus.

Remdamasis „Canva Marketing & AI Report 2025“, Aurimas pažymėjo, kad 94 % pasaulio rinkodaros vadovų jau turi suplanuotus AI biudžetus, o 78 % laiko AI ilgalaikės strategijos ašimi. Jis ragino Lietuvos įmones neatsilikti ir integruoti AI strategiškai.

Aurimas atkreipė dėmesį, kad 64 % rinkodaros specialistų susiduria su sunkumais renkantis tinkamus AI įrankius. Jis rekomendavo aiškiai apibrėžti problemą, įtraukti strategus ir vykdytojus bei atlikti mini testus prieš diegiant sprendimus.

Jis pateikė gaires, kaip pasirinkti tinkamiausią ChatGPT modelį: GPT-4o kasdienėms užduotims, o3 sudėtingiems iššūkiams. Taip pat paminėjo funkcijas kaip „Browsing“ ir „Canvas“, kurias padeda didinti produktyvumą.

Aurimo Pauliaus Girčio dalyvavimas AI WEEK Milan 2025 atspindi gilų įsitraukimą į dirbtinio intelekto sritį. Jo įžvalgos apie kompetencijų ugdymą, strateginę integraciją ir efektyvų įrankių naudojimą prisideda prie platesnio AI vaidmens versle ir rinkodaroje supratimo.

LinkedIn įrašai:

https://www.linkedin.com/posts/aurimasgircys_ai-week-turb%C5%ABt-did%C5%BEiaus-renginys-kuris-activity-7327955275376672769-swrT?utm_source=share&utm_medium=member_desktop&rcm=ACoAAAH0Rb4B7ArS4AJeGhMf7DJdjBRetecUITs

https://www.linkedin.com/posts/aurimasgircys_vyksta-registracija-%C4%AF-ai-week-activity-7327956037519429632-w1iV?utm_source=share&utm_medium=member_desktop&rcm=ACoAAAH0Rb4B7ArS4AJeGhMf7DJdjBRetecUITs

https://www.linkedin.com/posts/aurimasgircys_what-happens-to-ai-when-there-is-too-much-activity-7327962479542595584-Q7z1?utm_source=share&utm_medium=member_desktop&rcm=ACoAAAH0Rb4B7ArS4AJeGhMf7DJdjBRetecUITs

https://www.linkedin.com/posts/aurimasgircys_before-you-hire-your-next-employee-think-if-activity-7327963515208491009-VCZs?utm_source=share&utm_medium=member_desktop&rcm=ACoAAAH0Rb4B7ArS4AJeGhMf7DJdjBRetecUITs

https://www.linkedin.com/posts/aurimasgircys_some-people-say-that-its-not-about-prompting-activity-7327992188229828608-7JvG?utm_source=share&utm_medium=member_desktop&rcm=ACoAAAH0Rb4B7ArS4AJeGhMf7DJdjBRetecUITs

https://www.linkedin.com/posts/aurimasgircys_humans-can-fail-100-times-on-the-same-day-activity-7328023702070022145-109?utm_source=share&utm_medium=member_desktop&rcm=ACoAAAH0Rb4B7ArS4AJeGhMf7DJdjBRetecUITs

https://www.linkedin.com/posts/aurimasgircys_great-meeting-people-from-aimage-and-learning-activity-7328026879657951232-NyJx?utm_source=share&utm_medium=member_desktop&rcm=ACoAAAH0Rb4B7ArS4AJeGhMf7DJdjBRetecUITs

https://www.linkedin.com/posts/aurimasgircys_net-64-rinkodaros-specialist%C5%B3-susiduria-activity-7328038193570713602-S1Vo?utm_source=share&utm_medium=member_desktop&rcm=ACoAAAH0Rb4B7ArS4AJeGhMf7DJdjBRetecUITs

https://www.linkedin.com/posts/aurimasgircys_94-rinkodaros-vadov%C5%B3-pasaulyje-jau-turi-activity-7328040346779541504-7-0E?utm_source=share&utm_medium=member_desktop&rcm=ACoAAAH0Rb4B7ArS4AJeGhMf7DJdjBRetecUITs

https://www.linkedin.com/posts/aurimasgircys_kaip-per-30-sekund%C5%BEi%C5%B3-pasirinkti-tinkamiausi%C4%85-activity-7328043770833211395-WOOo?utm_source=share&utm_medium=member_desktop&rcm=ACoAAAH0Rb4B7ArS4AJeGhMf7DJdjBRetecUITs

https://www.linkedin.com/posts/aurimasgircys_tiek-daug-skirting%C5%B3-prane%C5%A1im%C5%B3-vienu-metu-activity-7328045285455060992-Lmpp?utm_source=share&utm_medium=member_desktop&rcm=ACoAAAH0Rb4B7ArS4AJeGhMf7DJdjBRetecUITs

https://www.linkedin.com/posts/aurimasgircys_antroji-ai-week-diena-prasid%C4%97jo-activity-7328322211814043649-8PIc?utm_source=share&utm_medium=member_desktop&rcm=ACoAAAH0Rb4B7ArS4AJeGhMf7DJdjBRetecUITs

https://www.linkedin.com/posts/aurimasgircys_kaip-para%C5%A1yti-ai-iliustracijas-u%C5%BEklaus%C4%85-activity-7328332227665149952-8XUe?utm_source=share&utm_medium=member_desktop&rcm=ACoAAAH0Rb4B7ArS4AJeGhMf7DJdjBRetecUITs

https://www.linkedin.com/posts/aurimasgircys_great-ai-output-starts-with-a-great-prompt-activity-7328334029768818688-iBCx?utm_source=share&utm_medium=member_desktop&rcm=ACoAAAH0Rb4B7ArS4AJeGhMf7DJdjBRetecUITs

https://www.linkedin.com/posts/aurimasgircys_94-of-workers-want-training-in-generative-activity-7328342847240269824-gL3?utm_source=share&utm_medium=member_desktop&rcm=ACoAAAH0Rb4B7ArS4AJeGhMf7DJdjBRetecUITs

Išvados

Hipotezių testų išvados: Remiantis atliktu tyrimu, galima prieiti prie šių išvadų dėl iškeltų hipotezių:

1. Dėl teksto kokybės (prasmingumo, nuoseklumo, tikslumo): *patvirtinta*. GPT-4 modeliu paremta sistema generuoja aukštos kokybės turinį – tekstai sklandūs, gerai struktūruoti, be akivaizdžių semantinių klaidų. Žmogaus vertintojai įvertino straipsnių kokybę beveik maksimaliais balais. Tai rodo, kad modelis pajėgus kurti prasmingus ir rišlius straipsnius įvairiomis temomis. Tik retais atvejais pastebėti nuoseklumo nukrypimai (pvz., minimalūs loginiai nesklandumai itin sudėtingose temose).
2. Dėl turinio originalumo: *patvirtinta*. Visi sugeneruoti straipsniai pasirodė esą originalūs pagal plagijavimo patikras. Tai reiškia, kad modelis formuluoja mintis savaip, nekopijuodamas ištisių fragmentų iš mokymo duomenų. Žinoma, tam tikros frazės ar sakinių struktūros gali kartotis, tačiau jos būdingos bendrinei kalbai ir nėra laikomos plagijuotomis. Ši išvada labai svarbi – ji suponuoja, kad GPT-4 generuojamas turinys tinkamas naudoti viešai autentiškumo požiūriu.⁵
3. Dėl aktualių temų identifikavimo: *iš esmės patvirtinta*. Modelis geba identifikuoti ir pabrėžti svarbiausius aspektus duota tema. Bandymuose, kai užklauso formuluotė buvo bendra, AI sugeneruotuose straipsniuose vis tiek atsirasdavo aktualiausi diskutuoti klausimai. Pavyzdžiui, pateikus temą „Klimato kaitos poveikis ekonomikai“, modelio straipsnyje aptarti pagrindiniai dalykai (žalojami sektoriai, prisitaikymo kaštai, tarptautinės politikos aspektai) – t. y. tie patys, kuriuos būtų paminėję ir žmogaus ekspertai. Trūkumas tas, kad modelio žinios ribotos mokymo duomenų aprėptimi – naujausių 2024–2025 m. įvykių ar tendencijų jis gali nežinoti. Pastebėjome, kad jei tema susijusi su labai nauju įvykiu, AI straipsnis apsiriboja bendra informacija arba istoriniais duomenimis. Tačiau tai nėra modelio „nesugebėjimas“ identifikuoti svarbios temos – tai veikiau žinių atnaujinimo klausimas. Suteikus aktualią informaciją kaip dalį užklauso, modelis puikiai ją įtraukia. Taigi, hipotezė, kad AI atskirs kas svarbu, pasitvirtino – su sąlyga, jog jis turi prieigą prie reikalingos informacijos.
4. Dėl stiliaus personalizavimo skirtingoms auditorijoms: *patvirtinta*. GPT-4 modelis demonstruoja lankstumą stilistiškai: gali rašyti formaliu, moksliniu tonu, gali kurti lengvabūdžiškus populiarius tekstus ar pardaviminiuosius rinkodaros straipsnius. Mūsų testai parodė, kad pakoregavus užklausa, tas pats modelis ženkliai pakeičia teksto registrą. Vertintojai aiškiai atskyrė tekstus skirtus skirtingoms auditorijoms, tad personalizavimo

⁵ Vis dėlto rekomenduojama ir toliau naudoti automatines patikros priemones kiekvienam publikuojamam straipsniui, ypač jei tema siaura ir yra rizika, kad modelis galėjo remtis vienu šaltiniu.

galimybė reali. Tai didelis privalumas, nes leidžia tą patį turinį pritaikyti skirtingiems kanalams ir skaitytojams automatiškai. Reikia pažymėti, kad modelis neturi vidinės „personos“ sąvokos – jis tiesiog vykdo vartotojo nurodymus. Todėl personalizacijos sėkmė priklauso nuo to, kaip aiškiai ir teisingai suformuluosime instrukcijas. Ateityje svarstyti įdiegti standartizuotus stiliaus nustatymus (pvz., pasirinkimus meniu: „rašyti akademinu stiliumi“, „rašyti jaunatvišku stiliumi“ ir pan.), kad vartotojui nereikėtų kiekvieną kartą ranka aprašyti pageidaujamo tono.

5. Dėl klaidingos informacijos ir šališkumo nebuvimo: *iš dalies patvirtinta*. Modelis akivaizdžiai sukurtas taip, kad vengia neetiško turinio – per visus testus nepateikė nieko, kas būtų netinkama ar tendencinga tyčia. Klaidingos informacijos klausimu: dauguma faktų teisingi, tačiau modelis nėra 100% neaktyvus – keletas klaidų pasitaikė. Tai rodo, kad visiškai pasikliauti be žmogaus patikros kol kas negalime (ypač kritinėse srityse). Šališkumo klausimu: akivaizdžios diskriminacijos nepastebėta, bet smulkūs šališkumai (pvz., lyčių stereotipų atspindžiai kalboje) egzistuoja, kad ir nedažnai. Ši hipotezė išpildyta tuo aspektu, kad AI stengiasi būti faktiškai tikslus ir nešališkas, tačiau absoliučios garantijos dar nėra. Mums teko įsitikinti, jog GPT-4 tiesiog yra daug atsargesnis ir tikslesnis nei ankstesnės kartos modeliai, bet vartotojas turėtų žinoti jo ribas. OpenAI tvirtina, kad „regurgitacija“ (mokymo duomenų kartojimas) – retas bugas, o šališkumai mažinami – mūsų tyrimas neprieštarauja tam, bet ir parodo, kad žmogiška priežiūra vis tiek pageidautina.

Rekomendacijos tolesnei plėtrai

Remiantis tyrimo rezultatais, pateikiamos kelios rekomendacijos, kaip būtų galima pagerinti AI straipsnių generavimo sistemą ir jos diegimą praktikoje:

- Faktų tikrinimas realiu laiku: Siekiant sumažinti likusių faktinių netikslumų skaičių, rekomenduojama integruoti automatinį faktų tikrinimo modulį. Pavyzdžiui, pasitelkti išorinius šaltinius (Vikipediją, duomenų bazines) – modelis sugeneravęs teiginį galėtų jį palyginti su patikimų šaltinių duomenimis. Kitas būdas – naudoti hibridinį metodą: GPT-4 generuoja turinį, o tuomet kitas modelis arba taisyklių rinkinys patikrina, ar jame nėra žinomų klaidų. Tai svarbu ypač naujienų ar mokslinėse temose, kur reikalingas didelis tikslumas.
- Reguliari modelio atnaujinimo strategija: Kadangi modelio žinios fiksuotos iki tam tikro laiko, būtina numatyti periodišką atnaujinimą arba integruoti mechanizmus, leidžiančius gauti informaciją apie naujausius įvykius. Pvz., galima naudoti OpenAI suteikiamą galimybę su GPT-4 naršyti internete (jei tokia prieiga yra) arba treniruoti atskirą papildinį

naujienu duomenimis. Taip sistema galės generuoti straipsnius ir visiškai naujomis temomis neprarasdama aktualumo.

- Stiliaus ir tono nustatymai vartotojui: Kaip minėta, personalizacijai būtų naudinga sukurti paprastus valdiklius. Siūloma sukurti keletą išankstinių stiliaus profilių (pvz., „Akademini“, „Naujienu pranešimas“, „Reklamini“, „Humoristini“ ir pan.) – vartotojas galėtų pasirinkti profilį prieš generuodamas tekstą. Tai užtikrintų dar nuoseklesnį stilistinį atitikimą. Galima remtis konkuruojančių įrankių praktika (Jasper leidžia nustatyti toną, auditoriją iš anksto). Toks funkcionalumas padidins sistemos patrauklumą praktiškai naudojant.
- Vartotojo sąsajos tobulinimas ir integracijos: Nors šiame tyrime prototipas buvo minimalus, galvojant apie produkto vystymą reikėtų atsižvelgti į naudotojų patirtį. Rekomenduojama integruoti sistemą į turinio valdymo platformas (pvz., WordPress įskiepiai), kad vartotojai galėtų generuoti straipsnius ten, kur juos publikuos. Tai leistų konkuruoti su kitais rinkos produktais, suteikiant daugiau vertės nei vien tik tekstas.
- Etikos gairės ir moderavimas: Rekomenduojama parengti aiškias naudojimo gaires galutiniams vartotojams, pabrėžiant, kad nors modelis stengiasi būti tikslus ir nešališkas, galutinė atsakomybė už turinio patikrą tenka žmogui. Sistemoje galima įdiegti įspėjimus, jei aptinkami potencialiai jautrūs teiginiai ar nepatikrinti faktai. Taip pat verta reguliariai atlikti auditus dėl šališkumo – pavyzdžiui, kas kelis mėnesius generuoti tekstus tam tikromis temomis ir tikrinti, ar neatsiranda nepageidaujamų tendencijų (tai ypač aktualu, jei modelis bus toliau finetuninamas su naujais duomenimis).
- Tęstinis mokymasis iš vartotojų grįžtamojo ryšio: Įdiegus sistemą, galima rinkti vartotojų atsiliepimus apie sugeneruotų straipsnių kokybę. Ypač vertingi būtų signalai, jei vartotojas redaguoja gautą tekstą – šiuos redagavimus galima analizuoti automatiškai ir identifikuoti dažniausiai taisomas vietas (pvz., gal modelis linkęs peržengti pageidaujamą toną ar trūksta tam tikros struktūros). Tai leistų nuolat gerinti modelio atsakymus pritaikant *reinforcement learning from human feedback* metodus ateityje.

Apibendrinimas

Atliktas tyrimas patvirtino, kad GPT-4 modeliu grįsta straipsnių generavimo sistema yra pajėgi generuoti kokybišką, originalų ir auditorijai pritaikomą turinį. Nustatyti trūkumai – nedidelis faktinių klaidų ir subtilių šališkumų procentas – yra valdoma rizika, kurią galima sumažinti techninėmis priemonėmis ir redakcinėmis praktikomis. Ši sistema turi potencialo padidinti turinio kūrimo efektyvumą žiniasklaidoje, rinkodaroje ir kitose srityse, tačiau rekomenduojama ją diegti laikantis principo „žmogus+AI“: t. y. AI generuoja juodraščių, o žmogus redaguoja ir

patvirtina galutinį variantą. Toks derinys užtikrins aukščiausią kokybę ir patikimumą. Tolesni žingsniai – technologiniai patobulinimai ir nuolatinis stebėjimas – padės šią inovatyvią sistemą paversti patikimu įrankiu kasdienėje praktikoje.