

# Matematikos Sprendimas

Dominykas Dulevičius

Informatikos institutas

Matematikos ir informatikos fakultetas  
Vilnius, Lietuva

dominykas.dulevicius@mif.stud.vu.lt

Arnas Bulka

Informatikos institutas

Matematikos ir informatikos fakultetas  
Vilnius, Lietuva

arnas.bulka@mif.stud.vu.lt

Pijus Kizerskis

Informatikos institutas

Matematikos ir informatikos fakultetas  
Vilnius, Lietuva

pijus.kizerskis@mif.stud.vu.lt

**Santrauka**—Projektinio darbo metu tyrėme ir lyginome internete paplitusių populiarių modelių galimybes ir tikslumą sprendžiant matematinius uždavinius bei lyginome rezultatus su TP-Transformer [7] tipo modeliu, kuris buvo specialiai aptreniruotas su DeepMind kūrėjų „Mathematics-Dataset“ duomenų rinkiniu.

**Raktažodžiai**—TP - Tenzoriaus-Produkto

## I. ĮVADAS

Matematinis samprotavimas – esminis žmogaus intelekto gebėjimas, veikiantis kaip pagrindinis mokslinės pažangos ir civilizacijos vystymosi jėga. Matematikos sprendimai nėra intuityvūs – mes uždavinius sprendžiame taikydami įvairias taisykles, interpretuodami simbolius bei taikydami globalias aksiomas. Pasauliui sparčiai technologiškai žengiant kartu su skaitmenizacija ir dirbtinio intelekto įtaka kasdieniniams darbams ir įpročiams, mūsų darbo tikslas ištirti kaip efektyviai skirtingi modeliai ir jų architektūros sprendžia įvairaus lygio ir skirtingų temų matematinius uždavinius.

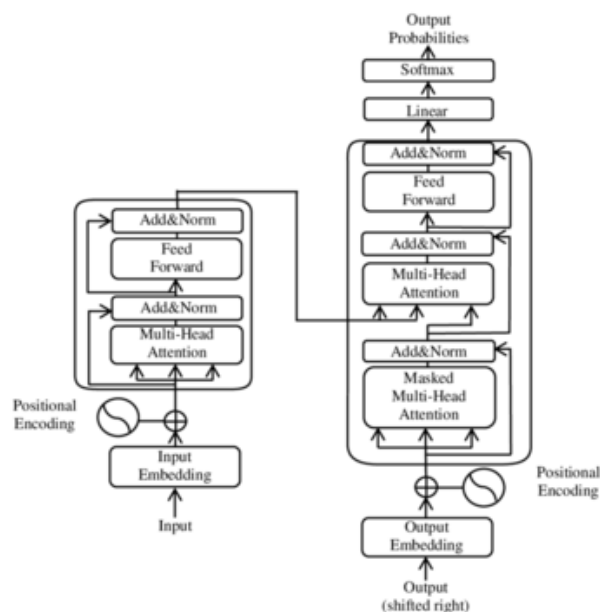
## II. DUOMENŲ RINKINYS „MATHEMATICS-DATASET“ [6]

Mūsų pasirinktas duomenų rinkinys yra sudarytas tik iš laisvos formos sintetiškai sugeneruotų matematinių uždavinių iš skirtingų temų. Duomenų rinkinyje uždaviniai yra tik tekstiniai, neprašoma nupiešti grafikų ar medžių, nes tokius duomenis būtų daug sudėtingiau validuoti ir modeliai taptų kur kas sudėtingesni. Uždavinių generatorius sukurtas akcentuojant, kad uždaviniai būtų kuo aktualesni ir atspindėtų kasdienio gyvenimo „matematikos egzaminą“. Šis duomenų rinkinys buvo sukurtas ištirti bei pagerinti neuroninių tinklų modelių matematinio samprotavimo galimybes. Plačiau žr. „Analysing mathematical reasoning of neural models“, [3]. Straipsnyje lygino LSTM (angl. *Long Short-Term Memory*), RMC (angl. *Relational Memory Core*) ir Transformer tipų modelių variacijas. Geriausią tikslumą parodė TP-Transformer [7] modelis kurį ir pasirinkome, kaip pagrindinį tyrimo objektą.

## III. TRANSFORMERIS

### A. Veikimas

Populiariausias bei plačiausiai paplitęs natūralios kalbos modelis - transformeris. [3] Jis pakeitė LSTM aplenkdamas pastarąjį tiek efektyvesniu apmokymu, tiek tikslesniais bei greitesniais atsakymais. LSTM modelis tekstą apdoroja eidamas per kiekvieną žodį iš eilės bandydamas „prisiminti“ visą sakinio esmę. Transformeris apdoroja kiek įmanoma daugiau



1 pav. – Transformerio architektūra

teksto vienu metu ir, atkreipdamas dėmesį į raktinius žodžius, bando suprasti teksto esmę.

### B. Moduliai

Iš pirmo žvilgsnio modelio architektūra gali pasirodyti sudėtinga „1 pav.“, tačiau ją galima išskirti į dvi pagrindines dalis:

- **Koduotuvas** (angl. *Encoder*) yra atsakingas už įeities duomenų susistemizavimą bei užkodavimą. Tai modeliui padeda suvienodinti gaunamus duomenis bei juos paruošti kitam transformerio moduliui - dekoderiui.
- **Dekoderis** (angl. *Decoder*) iš koduotuvo gauna galutinę paslėptą būseną pagal kurią sprendžia kiekvieno žodžio svarbą. Susidėliojęs konteksto vektorių dekoderis atiduoda galutinę išvestį.

### C. Dėmesys

Turbūt svarbiausias transformeryje naudojamas sluoksnis - susifokusavimas (angl. *self-attention*). Būtent dėl šios savybės transformeris geriau gali perprasti tekstą. Susifokusavimas figūruoja tiek koduotuvo, tiek dekoderio moduluose

padėdamas išgauti „paslėptas“ būsenas bei apskaičiuoti kiekvieno įvesto žodžio svarbumą.

#### IV. NAUDOJAMI APMOKYTI MODELIAI

Lyginti su TP-Transformer pasirinkome 5 iš anksto apmokytus klausimų-atsakymų modelius: **iAsk.AI**, **DeepAI**, **Bing AI**, **CHAT GPT-4**. Šių modelių tikslumą matematikos uždavinių sprendimui lyginsime su **ischlag/TP-Transformer** modeliu, specialiai pritaikytą matematikos uždaviniams spręsti.

##### A. ischlag/TP-Transformer

Šio modelio sukūrimui į standartinį transformerio tipo modelį buvo integruota Tenzoriaus-Produkto reprezentacija, tokiu būdu pagerinant išskirtinius sąryšius tarp transformerio vidinių struktūrų. [7] Modelis buvo apmokytas naudojant „Mathematics-Dataset“, kuris buvo išskaidytas į 56 skirtingas matematikos šakas, siekiant pagerinti modelio tikslumą bei atpažinamumą. Mūsų pasirinkta iš anksto išreniruota modelio versija buvo treniruojama su 1.7 milijono duomenų. Modelį galima išbandyti Google Colab platformoje.

##### B. iAsk.AI

Šis kalbos variklis naudoja panašias technologijas, kaip ir ChatGPT variklis, tačiau papildomas dėmesys yra skiriamas optimizuoti natūralios kalbos aprodorojimo modelį. [2] iAsk AI taip pat susideda iš specialiai pritaikyto, didelio masto Transformerio kalbos modelio. Šis modelis buvo išskirtinai mokomas remiantis patikimiausiais ir autoritetiškiausiais literatūros bei interneto šaltiniais, kas suteikia iAsk AI galimybę atsakyti į klausimus objektyviai, faktiškai ir be potencialaus subjektyvumo, kurio galėtų būti ChatGPT. Taip pat yra svarbu paminėti, kad gavus užduotį modelis pats sugeneruoja užklaudas interneto paieškos varikliams bei parenką tinkamiausią aplikaciją užduočiai įgyvendinti. Modelis yra pasiekiamas iAsk.AI puslapyje.

##### C. Bing AI

BingAI - Microsoft kompanijos sukurtas dirbtinio intelekto variklis, naudojamas Bing paieškos sistemoje. [4] Šis variklis naudoja tokias technologijas kaip gilųjį, mašininį mokymą ir natūralios kalbos apdorojimą, siekiant pagerinti paieškos rezultatus ir suteikti geriausius atsakymus vartotojams. Bing AI - nuolat tobulinamas ir mokomas naujų uždavinių, kad būtų užtikrinta patikima ir tiksli paieška. Taip pat Bing AI taikoma kitose Microsoft produktuose, tokiuose kaip Cortana ir Microsoft 365. Modelis yra pasiekiamas Bing AI paieškos variklyje.

##### D. DeepAI

DeepAI - GPT-2 modeliu paremtas natūralios kalbos modelis. [1] Šio, kaip ir kitų kalbos modelių, architektūroje slepiasi su dideliu kiekiu duomenų apmokytas transformeris. Modelis yra pasiekiamas Deep.AI puslapyje.

##### E. ChatGPT-4

ChatGPT-4 yra paremtas GPT modeliu, kuris naudoja Transformer architektūros dekodery. [5] Transformer architektūra turi koduotuvą ir dekodery, tačiau GPT naudoja tik autoregresinės formos dekodery, tai reiškia, kad jis yra optimizuotas tiksliai numatyti sekantį žodį sekoje. ChatGPT išskirtumas – duomenų rinkinio masyvumas. Jo duomenų rinkinys sudaro šimtai gigabaitų teksto. Modelis yra pasiekiamas Open AI puslapyje.

#### V. TYRIMO EIGA

Iš „Mathematics-Dataset“ duomenų rinkinio 56 matematinių uždavinių kategorijų išsirinkome po dvi klausimų-atsakymų poras. [6] Iš viso gavosi duomenų rinkinys, su 112 klausimų ir 112 atsakymų. Tada visus klausimus užduosime 4 pasirinktiems klausimų-atsakymų modeliams bei ischlag/TP-Transformer [7] modeliui ir skaičiuosime tikslumą - kiek iš užduotų klausimų modeliai sugebėjo atsakyti teisingai.

Toliau, padarėme kitą duomenų rinkinį iš 56 matematinių klausimų tokiu pačiu formatu ir iš tokių pačių temų, tačiau šie klausimai jau nėra iš „Mathematics-Dataset“ duomenų rinkinio, o tai yra atsitiktiniai klausimai. Šiems klausimams apskaičiavome atsakymus ir šiuos klausimus taip pat paduosime visiems 5 modeliams ir žiūrėsime, kaip skirsis atsakymų tikslumas. Ypač įdomu turėtų būti stebėti ischlag/TP-Transformer modelio tikslumą ant duomenų, kurių nebuvo jo treniruojamoje duomenų aibėje.

Klausimų ir atsakymų pavyzdžiai iš kiekvienos temos:

##### A. Algebra

- Klausimas: Suppose  $-2*i + 20 = 5*o$ ,  $2*o - 4 = 3*i + o$ . Suppose  $-g + 2 = i$ ,  $4*g - 6*g = -5*u - 19$ . Let  $f$  be  $-6 + 3 + 2 + u/(-1)$ . Solve  $-f*l = -1$  for  $l$ .
- Atsakymas: 0

##### B. Aritmetika

- Klausimas: What is the tenth root of 95210445 to the nearest integer?
- Atsakymas: 6

##### C. Matematinė Analizė

- Klausimas: Let  $s(w)$  be the third derivative of  $-7*w^{**2} - 1/30*w^{**5} - 23/6*w^{**3} + 0 + 0*w^{**6} + w - 1/42*w^{**7} + 1/24*w^{**4}$ . What is the second derivative of  $s(l)$  wrt  $l$ ?
- Atsakymas:  $-60*l^{**2} - 4$

##### D. Palyginimas

- Klausimas: Suppose  $-15*p + 48 + 27 = 0$ . Suppose  $13 = 5*f - p*n - 2$ ,  $0 = 3*f - 2*n - 8$ . Let  $d$  be  $(-3 + 110/36)*f$ . Which is the third biggest value? (a)  $1/6$  (b)  $d$  (c)  $-4$
- Atsakymas:  $c$

##### E. Konvertavimas

- Klausimas: How many micrometers are there in seven halves of a millimeter?
- Atsakymas: 3500

### F. Skaičių Operacijos

- Klausimas: Let  $a(s) = -154s + 138$ . Let  $x$  be  $a(0)$ . Let  $r(f) = -f - 3$ . Let  $z$  be  $r(-5)$ . Suppose  $4v + u + u - 282 = 0$ ,  $-2v + z \cdot u = -x$ . What are the prime factors of  $v$ ?
- Atsakymas: 2, 5, 7

### G. Daugianariai

- Klausimas: Simplify  $((y/y^{1/4})^{3/4})^{4/7} \cdot (y^{3/4} \cdot y^{4/7})^{5/4} / (y^{2/3})^{4/3} \cdot (y^{2/3})^{4/3}$  assuming  $y$  is positive.
- Atsakymas:  $y^{(-30203/3948)}$

### H. Tikimybės

- Klausimas: Four letters picked without replacement from a: 1, b: 1, s: 1, r: 1, q: 2, f: 2. Give prob of sequence rbaq.
- Atsakymas: 1/840

Modelio spėjimai bus klasifikuojami kaip binariiniai: arba modelis atspėjo atsakymą, arba ne. Atsakymas bus laikomas teisingu jei atsakymo reikšmė yra tokia pati, nepriklausomai nuo atsakymo duomenų formato. Kaip vienas iš pavyzdžių galėtų būti: jeigu tikėtas atsakymas buvo 1.5, bet modelis grąžino 3/2, atsakymas buvo užskaitytas kaip teisingas. Tikslumo metrika bus skaičiuojama pagal tikslumo formulę [2 pav.]. Atsakymų rinkimo pavyzdinę lentelę galite matyti „I lentelė“.

$$\text{Tikslumas} = \frac{\text{Teisingi spėjimai}}{\text{Uždavinių skaičius}} \quad (1)$$

2 pav. – Tikslumo formulė

I lentelė – Chat GPT-4 atsakymų lentelė.

Klausimas	Tema	Spėjimas	Tiesa
Solve $12s + 5 = 65$ for $s$	Algebra	5	5
Solve $5a + 2.5 = 10$ for $a$	Algebra	3/2	1.5
...	...	...	...

## VI. REZULTATAI

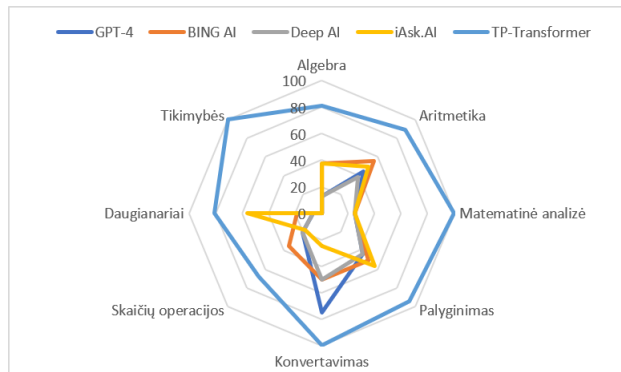
II lentelė – Klausimai iš „Mathematics-Dataset“.

Modelis	Teisingi atsakymai	Tikslumas %
ischlag/TP-Transformer	92	82.14
Bing	42	37.5
iAsk.Ai	41	36.61
Chat GPT-4	30	26.78
DeepAI	27	24.11

Sudarėme dvi lenteles bei diagramą rezultatų analizėms. „II lentelė“ parodo, kiek klausimų iš 112 pateiktų teisingai atspėjo modeliai iš „Mathematics-Dataset“ [6] duomenų rinkinio. „III lentelė“ atvaizduoja, kaip sekėsi tiems patiems modeliams spręsti kitus, atsitiktinius 56 klausimus iš tokių pačių matematinių temų.

III lentelė – Atsitiktiniai Klausimai

Modelis	Teisingi atsakymai	Tikslumas %
iAsk.Ai	35	62.5
Bing	29	51.79
DeepAI	26	46.43
ischlag/TP-Transformer	20	36.71
Chat GPT-4	18	32.14



3 pav. – Modelių tikslumai kiekvienai matematinei šakai

Toliau, kiekvienai matematikos šakai: Algebra, Aritmetika, Matematinė analizė, Palyginimas, Konvertavimas, Skaičių operacijos, Daugianariai, Tikimybės, lyginome modelių tikslumą panaudodami klausimus iš „Mathematics-Dataset“. Algebros šaką sudarė 16 klausimų, Aritmetikos - 18, Matematinės analizės - 4, Palyginimo - 16, Konvertavimo - 4, Skaičių operacijų - 34, Daugianarių - 16, Tikimybės - 4. Rezultatus, kaip tiksliai modeliai sekėsi atspėti klausimus, galime matyti „3 pav.“

## VII. IŠVADOS

- Geriausias modelis „Mathematics-Dataset“ duomenų rinkiniui yra ischlag/TP-Transformer su 82.15% tikslumu.
- Modelis ischlag/TP-Transformer geriausiai sprendžia uždavinius visoms matematinėms šakoms iš „Mathematics-Dataset“ duomenų rinkinio.
- Iš „3 pav.“ galime pastebėti, kad blogiausiai visi modeliai sprendžia tikimybių uždavinius. Tai reiškia, kad treniruojant modelius spręsti matematinius uždavinius daugiau dėmesio turėtų būti skirta tikimybių uždaviniams.
- Geriausiai pasirodęs modelis su atsitiktiniais klausimais yra iAsk.Ai surinkęs 62.5%.
- Galime matyti, kad atsitiktinius klausimus tiksliausiai sprendė iAsk.Ai modelis todėl, nes jo veikimo principas yra paremtas interneto paieškos tinkamiausiai aplikacijai spręsti užduotą uždavinį. Tą geriausiai iliustruoja, kad didžiajai daliai klausimų iAsk.Ai naudojo WolframAlpha[šaltinis] matematikos variklį.
- Mažiausią tikslumą su atsitiktiniais klausimais surinko Chat GPT-4, nes, mūsų manymu, šis modelis yra labiau pritaikytas teksto generavimui, nei specifinių užduočių sprendimui.
- Iš ischlag/TP-Transformer tikslumo atsitiktiniams klausimams galime pastebėti, kad šis modelis gali gerai spręsti

tik tuos klausimus kurie buvo „Mathematics-Dataset“ duomenų rinkinyje, nes būtent ant šių duomenų jis buvo ištreniruotas.

- Nesvarbu koks geras ir su kiek daug duomenų bus apmokytas modelis (pavyzdžiui ChatGPT-4), jeigu jis nebuvo treniruotas su specifinės srities duomenimis, jis gerai toje srityje nepasirodys.

#### VIII. REZULTATŲ TAIKYMO IR TOBULINIMO GALIMYBĖS

- Pabandyti papildomai su „Mathematics-Dataset“ ištreniruoti (angl. *fine-tune*) jau egzistuojančius transformerių modelius ir palyginti, kaip šis duomenų rinkinys pagerina matematinių uždavinių sprendimą.
- Palyginti LSTM, Scaling-Transformer bei Graphic Logic Reasoner modelių tikslumą sprendžiant matematinius uždavinius [6]
- Pasirinkti kelis į matematiką orientuotus duomenų rinkinius, ne tik „Mathematics-Dataset“, su jais papildomai ištreniruoti tą patį pasirinktą modelį ir palyginti kaip skirtingai jie pagerina uždavinių sprendimą.

#### LITERATŪRA

- [1] Deepai. <https://www.deep.ai>, 2019.
- [2] AI Search Inc. <https://www.airesearchinc.com>, 2022.
- [3] Felix Hill David Saxton, Edward Grefenstette and Pushmeet Kohli. Analysing mathematical reasoning abilities of neural models. *ICLR 2019*, 2019.
- [4] Yusuf Mehdi. Announcing the next wave of ai innovation with microsoft bing and edge. Blog post, May 2023.
- [5] OpenAI. Chatgpt: Large-scale language models for conversational ai. *OpenAI Blog*, 2023.
- [6] David Saxton, Edward Grefenstette, Felix Hill, and Pushmeet Kohli. Mathematics dataset. Dataset, 2020.
- [7] Imanol Schlag, Paul Smolensky, Roland Fernandez, Nebojsa Jojic, Jürgen Schmidhuber, and Jianfeng Gao. Enhancing the transformer with explicit relational encoding for math problem solving. *arXiv preprint arXiv:1910.06611*, 2019.