TABIIY TILNI QAYTA ISHLASH (NLP) TEXNOLOGIYALARINING ZAMONAVIY RIVOJLANISH TENDENSIYALARI

Qarshiboyev Vosid Vaxob o'g'li
Toshkent davlat iqtisodiyot universiteti, talaba
e-mail: qvosid@gmail.com
+998 99 855 99 71
Doi: https://doi.org/10.5281/zenodo.20285782

Annotatsiya: Mazkur maqolada tabiiy tilni qayta ishlash (Natural Language Processing — NLP) texnologiyalarining zamonaviy rivojlanish tendensiyalari, ularning sun'iy intellekt tizimlaridagi o'rni hamda amaliy qo'llanish yo'nalishlari ilmiy manbalar asosida tahlil qilinadi. NLP texnologiyalarining tarixiy shakllanishi, neyron tarmoqlar va transformer arxitekturasining rivojlanishi, katta til modellari (Large Language Models — LLM) paydo bo'lishi va ularning samaradorligi yoritilgan. Shuningdek, mashinaviy tarjima, matn generatsiyasi, sentiment tahlili, savol-javob tizimlari hamda multimodal sun'iy intellekt tizimlaridagi yangiliklar ko'rib chiqilgan. Tadqiqot davomida BERT, GPT, T5, PaLM va boshqa zamonaviy modellar faoliyati ilmiy manbalar asosida tahlil qilinib, NLP texnologiyalarining istiqboldagi rivojlanish yo'nalishlari aniqlangan. Tadqiqot natijalari NLP sohasida transformer modellarining dominant texnologiyaga aylanganini hamda katta hajmdagi ma'lumotlar va hisoblash quvvatlari ushbu rivojlanishning asosiy omili ekanligini ko'rsatadi.

Kalit so'zlar: Tabiiy tilni qayta ishlash, NLP, sun'iy intellekt, transformer, katta til modellari, BERT, GPT, mashinaviy tarjima, neyron tarmoqlar, multimodal tizimlar.

Аннотация: В данной статье на основе научных источников анализируются современные тенденции развития технологий обработки естественного языка (Natural Language Processing — NLP), их роль в системах искусственного интеллекта, а также направления практического применения. Освещаются историческое становление NLP-технологий, развитие нейронных сетей и архитектуры transformer, появление больших языковых моделей (Large Language Models — LLM) и их эффективность. Кроме того, рассматриваются новшества в области машинного перевода, генерации текста, анализа тональности, систем вопросов и ответов, а также мультимодальных систем искусственного интеллекта. В ходе исследования на основе научных источников были проанализированы модели BERT, GPT, T5, PaLM и другие современные модели, а также определены перспективные направления развития NLP-технологий. Результаты исследования показывают, что transformer-модели стали доминирующей технологией в области NLP, а основными факторами данного развития являются большие объемы данных и вычислительные мощности.

Ключевые слова: обработка естественного языка, NLP, искусственный интеллект, transformer, большие языковые модели, BERT, GPT, машинный перевод, нейронные сети, мультимодальные системы.

Abstract: This article analyzes the modern development trends of Natural Language Processing (NLP) technologies, their role in artificial intelligence systems, and their practical application areas based on scientific sources. The historical formation of NLP technologies, the development of neural networks and transformer architecture, the emergence of Large Language Models (LLMs), and their effectiveness are highlighted. In addition, recent advances in machine translation, text generation, sentiment analysis, question-answering systems, and multimodal

artificial intelligence systems are examined. During the research, the performance of BERT, GPT, T5, PaLM, and other modern models was analyzed based on scientific sources, and the future development directions of NLP technologies were identified. The research results demonstrate that transformer models have become the dominant technology in the field of NLP, while large-scale data and computational power are the main driving factors behind this development.

Keywords: Natural Language Processing, NLP, artificial intelligence, transformer, large language models, BERT, GPT, machine translation, neural networks, multimodal systems.

Kirish

Tabiiy tilni qayta ishlash texnologiyalari inson va kompyuter o‘rtasidagi muloqotni avtomatlashtirishga xizmat qiluvchi sun’iy intellekt yo‘nalishlaridan biri hisoblanadi. NLP texnologiyalarining asosiy maqsadi inson tilini kompyuter tomonidan tushunish, qayta ishlash va generatsiya qilish jarayonlarini amalga oshirishdan iborat. So‘nggi yillarda katta hajmdagi ma’lumotlar, yuqori unumdor grafik protsessorlar (GPU) va chuqur o‘rganish algoritmlarining rivojlanishi NLP tizimlarining samaradorligini sezilarli darajada oshirdi [3].

Dastlabki NLP tizimlari qoidalarga asoslangan yondashuvlardan foydalangan bo‘lsa, keyinchalik statistik modellar keng qo‘llanila boshlandi. XXI asrning ikkinchi o‘n yilligidan boshlab esa chuqur neyron tarmoqlar va transformer arxitekturasi NLP texnologiyalarida tub burilish yasadi. 2017-yilda Google kompaniyasi tomonidan taklif etilgan transformer modeli NLP sohasida yangi bosqichni boshlab berdi [1]. Ushbu model rekurrent neyron tarmoqlariga nisbatan yuqori aniqlik va tezlikni ta’minladi.

Bugungi kunda NLP texnologiyalari qidiruv tizimlari, virtual yordamchilar, avtomatik tarjima, tibbiy diagnostika, moliyaviy tahlil va ta’lim tizimlarida keng qo‘llanmoqda. Statista ma’lumotlariga ko‘ra, global NLP bozori hajmi 2021-yilda 11 milliard AQSh dollaridan oshgan bo‘lib, 2030-yilgacha ushbu ko‘rsatkich bir necha barobar o‘sishi prognoz qilinmoqda [4].

Metodologiya

Mazkur tadqiqotda ilmiy maqolalar, konferensiya materiallari, xalqaro tashkilotlar hisobotlari va yetakchi texnologik kompaniyalar tomonidan e’lon qilingan texnik hujjatlar tahlil qilindi. Tadqiqotning metodologik asosini quyidagi usullar tashkil etdi:

- ilmiy adabiyotlarni qiyosiy tahlil qilish;
- NLP modellarining samaradorligini solishtirish;
- statistik ma’lumotlarni tahlil qilish;
- zamonaviy NLP texnologiyalarining amaliy qo‘llanishini o‘rganish.

Tadqiqot davomida transformer arxitekturasi asosidagi BERT, GPT, T5 va PaLM modellarining ishlash prinsiplari hamda ularning tabiiy tilni qayta ishlashdagi samaradorligi o‘rganildi. Shuningdek, OpenAI, Google DeepMind va Meta kompaniyalari tomonidan taqdim etilgan ilmiy natijalar tahlil qilindi [5], [6].

Natijalar

Tadqiqot natijalari shuni ko‘rsatdiki, transformer arxitekturasi NLP texnologiyalarining rivojlanishida asosiy texnologik yechimga aylangan. Transformer modeli “attention mechanism” mexanizmi orqali matndagi kontekstual bog‘liqliklarni samarali aniqlaydi [1].

BERT modeli 2018-yilda Google tomonidan taqdim etilgan bo‘lib, ikki tomonlama kontekstni hisobga olish imkoniyati bilan NLP tizimlari aniqligini sezilarli oshirdi [2]. Ushbu model savol-javob tizimlari va matn klassifikatsiyasida yuqori natijalarga erishgan.

GPT oilasiga mansub modellar esa generativ sun’iy intellekt rivojlanishining asosiy omillaridan biri bo‘ldi. GPT-3 modeli 175 milliard parametrga ega bo‘lib, matn generatsiyasi, kod yozish, tarjima va muloqot tizimlarida yuqori samaradorlikni namoyon etdi [5]. Keyingi avlod modellari yanada katta parametrlar va multimodal imkoniyatlarga ega bo‘ldi.

Mashinaviy tarjima tizimlarida ham sezilarli o‘zgarishlar kuzatildi. Google Translate va DeepL kabi tizimlar transformer arxitekturasi yordamida tarjima sifatini ancha yaxshiladi. BLEU ko‘rsatkichlari asosida olib borilgan tadqiqotlar neyron tarjima tizimlari statistik tarjima modellariga qaraganda ancha samarali ekanligini ko‘rsatdi [7].

Sentiment tahlili, matn summarizatsiyasi va savol-javob tizimlarida ham NLP modellarining aniqligi oshgan. Stanford Question Answering Dataset (SQuAD) natijalariga ko‘ra, zamonaviy transformer modellarining ayrimlari inson natijalariga yaqin ko‘rsatkichlarga erishgan [8].

So‘nggi yillarda multimodal NLP tizimlari rivojlanmoqda. Ushbu tizimlar matn, rasm, audio va video ma’lumotlarini birgalikda qayta ishlash imkonini beradi. Google Gemini va GPT-4 kabi tizimlar multimodal sun’iy intellektning amaliy misollari hisoblanadi [9].

Tahlil va muhokama

NLP texnologiyalarining zamonaviy rivojlanish tendensiyalarini tahlil qilish shuni ko‘rsatadiki, asosiy e’tibor katta til modellari va generativ sun’iy intellekt tizimlariga qaratilgan. Katta hajmdagi ma’lumotlar bazalari va yuqori hisoblash quvvatlari modellar samaradorligini oshirishda muhim rol o‘ynamoqda [10].

Transformer arxitekturasi rekurrent neyron tarmoqlari bilan taqqoslaganda bir nechta ustunliklarga ega. Birinchidan, parallel hisoblash imkoniyati tufayli modelni o‘qitish tezligi oshadi. Ikkinchidan, uzoq kontekstual bog‘liqliklarni aniqlash samaradorligi yuqori bo‘ladi. Shu sababli transformer NLP tizimlarida asosiy standartga aylangan [1].

Biroq zamonaviy NLP tizimlari bilan bog‘liq ayrim muammolar ham mavjud. Eng muhim muammolardan biri katta hajmdagi hisoblash resurslariga ehtiyojdir. Masalan, GPT-3 modelini o‘qitish uchun minglab GPU qurilmalari va katta miqdorda elektr energiyasi talab etilgan [5]. Bu esa ekologik va iqtisodiy masalalarni keltirib chiqarmoqda.

Yana bir muhim muammo — modellar tomonidan noto‘g‘ri yoki yolg‘on ma’lumot generatsiya qilinishi hisoblanadi. Tadqiqotchilar ushbu holatni “hallucination” deb atashadi. Generativ modellar ayrim hollarda faktik asosga ega bo‘lmagan javoblarni ishlab chiqishi mumkin [11]. Shu sababli NLP tizimlarida faktlarni tekshirish mexanizmlarini rivojlantirish muhim hisoblanadi.

NLP texnologiyalarining til xilma-xilligiga ta’siri ham dolzarb masala hisoblanadi. Zamonaviy modellar asosan ingliz tilidagi katta korpuslarda o‘qitilganligi sababli resurslari kam bo‘lgan tillarda samaradorlik pastroq bo‘lishi mumkin [12]. Shu bois ko‘plab davlatlarda milliy tillar uchun NLP modellarini yaratish bo‘yicha tadqiqotlar olib borilmoqda.

O‘zbek tili uchun ham NLP texnologiyalarini rivojlantirish istiqbolli yo‘nalishlardan biridir. So‘nggi yillarda o‘zbek tilidagi matn korpuslarini shakllantirish, morfologik analizatorlar yaratish va avtomatik tarjima tizimlarini takomillashtirish bo‘yicha ilmiy ishlar amalga oshirilmoqda. Biroq o‘zbek tilidagi ochiq ma’lumotlar bazalari hajmi hali ham cheklangan.

NLP texnologiyalarining amaliy ahamiyati turli sohalarda ortib bormoqda. Tibbiyotda klinik ma’lumotlarni tahlil qilish, huquq sohasida hujjatlarni avtomatik qayta ishlash, ta’lim tizimida aqlli tutor tizimlarini yaratishda NLP texnologiyalaridan keng foydalanilmoqda [3].

Kelajakda NLP texnologiyalarining rivojlanishi multimodal sun’iy intellekt, energiya samarador modellar va etik sun’iy intellekt konsepsiyalari bilan bog‘liq bo‘lishi kutilmoqda. Shuningdek, kichik hajmdagi, ammo yuqori samaradorlikka ega modellar yaratish ustida ham faol tadqiqotlar olib borilmoqda.

Xulosa

Tabiiy tilni qayta ishlash texnologiyalari sun’iy intellektning eng tez rivojlanayotgan yo‘nalishlaridan biri hisoblanadi. Transformer arxitekturasi va katta til modellari NLP tizimlarining aniqligi va funksional imkoniyatlarini sezilarli darajada oshirdi. BERT, GPT, T5 va boshqa zamonaviy modellar mashinaviy tarjima, matn generatsiyasi, savol-javob tizimlari va multimodal sun’iy intellekt rivojlanishida muhim rol o‘ynamoqda.

Tadqiqot natijalari NLP texnologiyalarining amaliy ahamiyati ortib borayotganini ko‘rsatdi. Shu bilan birga, hisoblash resurslari talabi, etik muammolar va noto‘g‘ri ma’lumot generatsiyasi kabi masalalar hali ham dolzarb bo‘lib qolmoqda. Kelgusida energiya samarador va ko‘p tilli NLP tizimlarini yaratish ushbu sohaning asosiy rivojlanish yo‘nalishlaridan biri bo‘lishi kutilmoqda.

Foydalanilgan adabiyotlar ro‘yxati

1. Vaswani A., Shazeer N., Parmar N. et al. *Attention Is All You Need*. — 2017. — pp. 5998–6008.
2. Devlin J., Chang M.W., Lee K., Toutanova K. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. — 2018. — pp. 4171–4186.
3. Jurafsky D., Martin J.H. *Speech and Language Processing*. 3-nashr. — Pearson, 2023. — pp. 35–120.
4. Statista Research Department. *Natural Language Processing Market Size Worldwide*. — 2023. — pp. 1–5.
5. Brown T., Mann B., Ryder N. et al. *Language Models are Few-Shot Learners*. — 2020. — pp. 1877–1901.
6. Chowdhery A., Narang S., Devlin J. et al. *PaLM: Scaling Language Modeling with Pathways*. — 2022. — pp. 1–113.
7. Bahdanau D., Cho K., Bengio Y. *Neural Machine Translation by Jointly Learning to Align and Translate*. — 2015. — pp. 1–15.
8. Rajpurkar P., Zhang J., Lopyrev K., Liang P. *SQuAD: 100,000+ Questions for Machine Comprehension of Text*. — 2016. — pp. 2383–2392.
9. OpenAI. *GPT-4 Technical Report*. — 2023. — pp. 1–100.
10. Bommasani R., Hudson D.A., Adeli E. et al. *On the Opportunities and Risks of Foundation Models*. — Stanford University, 2021. — pp. 1–218.
11. Ji Z., Lee N., Frieske R. et al. *Survey of Hallucination in Natural Language Generation*. — 2022. — pp. 1–38.
12. Bender E.M., Gebru T., McMillan-Major A., Shmitchell S. *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* — 2021. — pp. 610–623.