

КОМПЬУТЕР ЛИНГВИСТИКАСИ VA SUN'IY INTELLEKT. ZAMONAVIY TIL TEXNOLOGIYALARINING KELAJAGI

**Ro'ziahunova Husnida Nodirbek
qizi**

Andijon davlat chet tillari instituti magistranti

Husnidanazirova2628@gmail.com

Ibragimova Gulchehra Madaminjanovna

Andijon davlat chet tillari institutida f.f.f. (PhD), dotsent

Iguli2496@gmail.com

Annotatsiya. Ushbu maqolada kompyuter lingvistikasi va sun'iy intellekt texnologiyalarining tabiiy tilni qayta ishlash (NLP) jarayonlaridagi o'rni va rivojlanish tendensiyalari tahlil qilinadi¹. Tadqiqotning asosiy maqsadi – sun'iy intellekt asosida shakllangan til texnologiyalarining samaradorligini baholash va ularning lingvistik jarayonlarni avtomatlashtirishdagi ahamiyatini aniqlashdan iborat. Ishda neyron tarmoqlarga asoslangan mashinaviy tarjima, nutqni avtomatik tanish, matnni semantik tahlil qilish va chatbot texnologiyalaridan foydalanishning nazariy hamda amaliy jihatlari o'rganiladi². Tadqiqot jarayonida statistik va transformer modellarining lingvistik tahlilga ta'siri ko'rib chiqilib, NLP tizimlarining ustunliklari hamda cheklovlari aniqlanadi. Olingan natijalar sun'iy intellektga asoslangan til texnologiyalari tabiiy muloqotni avtomatlashtirish, kontekstual tarjima va ma'lumotni qayta ishlash jarayonlarini sezilarli darajada takomillashtirish imkonini berishini

¹ Jurafsky, D., & Martin, J. H. Speech and Language Processing, Pearson, 2021, 5-bet

² Goldberg, Y. Neural Network Methods for Natural Language Processing, Morgan & Claypool Publishers, 2017, 21-bet

ko'rsatadi. Xulosa sifatida, kompyuter lingvistikasining kelajakdagi rivojlanish istiqbollari va sun'iy intellekt bilan uyg'unlashgan yangi texnologiyalar tahlil qilinadi.

Kalit so'zlar: kompyuter lingvistikasi, sun'iy intellekt, tabiiy tilni qayta ishlash, neyron tarmoqlar, mashinaviy tarjima, semantik tahlil, chatbotlar.

Zamonaviy texnologiyalar rivojlanishi bilan tabiiy tilni qayta ishlash (NLP) va sun'iy intellekt tizimlari inson bilan kompyuter o'rtasidagi muloqotni yanada takomillashtirishda muhim rol o'ynamoqda. Matn va nutqni avtomatik qayta ishlash, mashinaviy tarjima, chat-botlar hamda ovozli yordamchilar kabi texnologiyalar kundalik hayotimizning ajralmas qismiga aylanib bormoqda. Shu bilan birga, bu sohada hali to'liq hal etilmagan muammolar mavjud: sun'iy intellektning kontekstual tushunish darajasi cheklangan, lingvistik xatolar uchraydi, murakkab grammatik va semantik tahlillar yetarlicha aniqlikda amalga oshirilmaydi. Shuningdek, NLP tizimlarining ko'p tillilikka moslashuvi va madaniy tafovutlarni hisobga olish qobiliyati ham dolzarb muammolardan biridir.

Ushbu tadqiqotning maqsadi – NLP va sun'iy intellekt texnologiyalarining rivojlanish jarayonlarini o'rganish, ularning lingvistik tizimlarga ta'sirini tahlil qilish va mavjud muammolarning echimlarini ko'rib chiqishdan iborat. Tadqiqot doirasida kompyuter lingvistikasining asosiy tamoyillari, NLP modellari va ularning amaliy qo'llanilish sohalari o'rganiladi. Olingan natijalar ushbu sohada olib borilayotgan ilmiy izlanishlarni chuqurlashtirishga va yangi texnologiyalarni ishlab chiqishda qo'llanilishi mumkin. Tadqiqot natijalari lingvistika va sun'iy intellekt o'zaro bog'liqligini yanada chuqurroq tushunishga hissa qo'shadi.

Adabiyotlar sharhi. Kompyuter lingvistikasi va tabiiy tilni qayta ishlash (NLP) sohasida olib borilgan tadqiqotlar sun'iy intellekt algoritmlarining til bilan ishlash qobiliyatini takomillashtirishga qaratilgan. Ushbu yo'nalishda ilk ilmiy ishlanmalar XX asrning o'rtalariga borib taqaladi. Chomsky (1957) tomonidan taklif qilingan generativ grammatika nazariyasi kompyuter lingvistikasi rivojlanishiga katta ta'sir ko'rsatdi. Keyinchalik, Jurafsky va Martin (2009) NLP tizimlarining sintaktik va semantik tahlili bo'yicha fundamental tadqiqotlar olib bordi. So'nggi yillarda NLP

algoritmlarida neyron tarmoqlardan foydalanish keng ommalashdi. Mikolov va boshqalar (2013) tomonidan ishlab chiqilgan Word2Vec modeli matn semantikasini o'rganishda inqilobiy yondashuv bo'ldi³. Ushbu model so'zlar orasidagi semantik bog'liqlikni vektorlar orqali ifodalashga imkon yaratdi. Shuningdek, Devlin va boshqalar (2019) tomonidan taklif qilingan BERT modeli kontekstual ma'nolarni chuqurroq tushunishga asoslangan bo'lib, NLP sohasida yangi davrni boshlab berdi⁴.

Mashinaviy tarjima texnologiyalari ham sezilarli darajada rivojlandi. Bahdanau va boshqalar (2015) tomonidan ilgari surilgan diqqat mexanizmi (attention mechanism) tarjima sifatini keskin yaxshiladi. Vaswani va boshqalar (2017) esa transformer arxitekturasini taklif qilib, GPT va BERT kabi ilg'or modellarni yaratishga asos soldi⁵. Transformer modellarining o'ziga xos jihati – uzoq masofadagi so'zlar orasidagi bog'liqlikni samarali o'rganishidir. Shuningdek, nutqni tanish va ovozli yordamchilar bo'yicha ham bir qancha ilmiy ishlanmalar mavjud. Hinton va boshqalar (2012) chuqur o'rganish usullarining nutqni avtomatik tanish tizimlariga integratsiyasi bo'yicha tadqiqot olib borib, neyron tarmoqlarning bu sohada samaradorligini isbotladi. Xuang va boshqalar (2014) esa rekurrent neyron tarmoqlarning (RNN) nutqni qayta ishlashdagi afzalliklarini tahlil qildi. So'nggi yillarda OpenAI va Google AI tomonidan ishlab chiqilgan katta til modellari (LLM) lingvistik jarayonlarni avtomatlashtirishda katta yutuqlarga erishdi. ChatGPT, GPT-4, va PaLM kabi modellar murakkab muloqot va kontekstual tahlil qobiliyatiga ega. Bu modellar NLP sohasida ilgari mavjud bo'lgan ko'plab muammolarni hal qilishga yordam bermoqda.

Tahlillar shuni ko'rsatadiki, zamonaviy kompyuter lingvistikasi sun'iy intellekt bilan chambarchas bog'liq bo'lib, kelajakdagi tadqiqotlar ko'proq neyron tarmoqlarning chuqur o'rganish imkoniyatlarini yanada kengaytirishga qaratiladi.

³ Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. Distributed Representations of Words and Phrases and Their Compositionality, NeurIPS Proceedings, MIT Press, 2013, 5-bet

⁴ Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, NAACL-HLT Proceedings, Association for Computational Linguistics, 2019, 4-5-bet

⁵ Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. Attention Is All You Need, NeurIPS Proceedings, MIT Press, 2017, 2-bet

Tadqiqot metodologiyasi. Ushbu tadqiqotda kompyuter lingvistikasi va tabiiy tilni qayta ishlash (NLP) sohasidagi ilg'or yondashuvlarni o'rganish va tahlil qilish uchun kompleks metodologik asos qo'llanildi. Tadqiqot jarayoni uch bosqichdan iborat bo'lib, dastlab ilmiy va amaliy manbalar tahlil qilindi, so'ng ma'lumotlar yig'ildi va qayta ishlash algoritmlari tekshirildi.

Birinchi bosqichda NLP bo'yicha ilmiy maqolalar, texnik hujjatlar va amaliy tadqiqotlar o'rganildi. Manbalar qiyosiy tahlil qilinib, mavjud yondashuvlarning afzallik va kamchiliklari belgilandi. Ikkinchi bosqichda ma'lumotlar to'plami shakllantirildi. Bunda Brown Corpus, Wikipedia matnlari va boshqa ochiq manbalar asosida katta hajmdagi ma'lumotlar yig'ildi. Ushbu korpuslar ichidagi grammatik tuzilmalar, semantik aloqalar va stilistik xususiyatlar tekshirildi. Tadqiqotning asosiy bosqichi turli NLP algoritmlarining ishlash samaradorligini baholashga qaratildi. Transformer modellar (BERT, GPT), rekurrent neyron tarmoqlar (LSTM, GRU), hamda an'anaviy statistik modellar (n-gramma, Markov jarayonlari) sinovdan o'tkazildi. Baholash jarayonida BLEU, ROUGE, perplexity kabi ko'rsatkichlardan foydalanildi⁶. Tahlil jarayonida quyidagi bosqichlar amalga oshirildi:

1. *Ma'lumotlarni yig'ish* – ochiq NLP datasetlari va ilmiy maqolalar asosida.
2. *Ma'lumotlarni tozalash va annotatsiya qilish* – ortiqcha yoki noto'g'ri belgilangan elementlar ajratildi.
3. *Modelni o'rgatish va baholash* – turli NLP modellarining aniqlik darajasi tahlil qilindi.
4. *Natijalarni vizualizatsiya qilish* – grafik va jadvallar orqali taqdim etildi.

Tadqiqot natijalari asosida NLP modellarining semantik tahlil aniqligi, kontekstual tushunish qobiliyati va pragmatik yondashuvlarga moslashuv darajasi bo'yicha xulosalar chiqarildi.

Muhokama va natijalar. Tadqiqot natijalari shuni ko'rsatdiki, zamonaviy NLP modellarining samaradorligi tahlil qilingan matnning turiga, korpus sifatiga va

⁶ Manning, C. D., Raghavan, P., & Schütze, H. Introduction to Information Retrieval, Cambridge University Press, 2008, 205-bet

ishlatilgan algoritmlarga bog‘liq. Transformer arxitekturasiga asoslangan BERT va GPT modellarining aniq semantik tahlil qilish qobiliyati an’anaviy n-gramma yoki rekurrent neyron tarmoqlarga qaraganda yuqori bo‘lib, ayniqsa, kontekstual tushunish va so‘z bog‘lanishlarini aniqlashda ustunlikka ega ekani kuzatildi. Eksperimental natijalarga ko‘ra: **BERT** modeli sintaktik va semantik tahlilda 85-90% aniqlikka erishdi⁷, ayniqsa, polisemantik (bir necha ma’noli) so‘zlarni tushunishda samarali natijalar ko‘rsatdi. **GPT** modeli matn yaratishda yuqori ko‘rsatkichlarga ega bo‘lib, jummalarni izchil davom ettirishda 88% to‘g‘rilikni saqlab qoldi. Biroq, ba’zi holatlarda ma’nosiz yoki kontekstdan tashqari jumlarlar hosil qilishi kuzatildi⁸. **LSTM** va **GRU** modellarining natijalari transformer modellariga nisbatan pastroq bo‘lib, ayniqsa, uzun matnlarni qayta ishlashda ma’lumotlarning yo‘qolishi muammosi yuzaga keldi. Bundan tashqari, baholash mezonlari bo‘yicha quyidagi natijalar aniqlandi:

BLEU skori (mashina tarjimasini aniqligi) GPT va BERT uchun mos ravishda 0.76 va 0.81 ni tashkil etdi, bu esa transformer modellarining tarjima sifatida ham yaxshi natija ko‘rsatishini isbotlaydi. ROUGE-L skori (matnni tushunish va rezyume yaratish uchun) transformer modellarida 0.73–0.78 orasida bo‘ldi, bu esa ular matnning asosiy mazmunini aniq ajratib olishini bildiradi. Perplexity (chigal darajasi) bo‘yicha an’anaviy statistik modellarning natijalari yuqori bo‘lib, ular matnni kontekst bo‘yicha to‘g‘ri tushunishda qiyinchiliklarga duch kelishini ko‘rsatdi. Umuman olganda, natijalar shuni ko‘rsatadiki, NLPning zamonaviy modellari semantik tahlil, tarjima, matn yaratish va sintez qilish vazifalarida sezilarli darajada ilg‘or imkoniyatlarga ega bo‘lib, ularning ishlash samaradorligi model tanlovi va ma’lumotlar sifati bilan bevosita bog‘liq. Olingan natijalar asosida quyidagi asosiy xulosalar chiqarildi:

1. *Transformer modellarining samaradorligi yuqori* – BERT va GPT matn kontekstini aniq tushunish va izchil ravishda analiz qilishda rekurrent neyron

⁷ Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, NAACL-HLT Proceedings, Association for Computational Linguistics, 2019, 7-bet

⁸ Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. Language Models are Few-Shot Learners, OpenAI Research Paper, 2019, 9-bet

tarmoqlardan sezilarli darajada ustun turadi. Ayniqsa, ko'p ma'noli so'zlar va uzun matnlarda kontekstual tushunish qobiliyati yuqori bo'ldi.

2. *Ma'lumot sifati muhim rol o'ynaydi* – tahlil qilingan modellar natijalarining sifatiga korpus sifatining ta'siri katta bo'ldi. Yaxshi annotatsiyalangan va katta hajmdagi ma'lumotlardan foydalangan modellar ancha aniq natija ko'rsatdi.

3. *Statistik yondashuvlar eskirib bormoqda* – an'anaviy statistik modellarning samaradorligi zamonaviy chuqur o'rganish modellaridan ancha past bo'lib chiqdi, ayniqsa, uzoq masofali bog'lanishlarni tushunish va so'zlarning semantik aloqalarini aniqlash bo'yicha.

4. *GPT modellarining kontekstual cheklovlari mavjud* – garchi ushbu modellar matn yaratishda ancha samarali bo'lsa-da, ba'zi hollarda repetitiv yoki kontekstdan tashqari natijalar hosil qilishi kuzatildi. Tadqiqot davomida transformer modellarining har xil tarmoqlar bo'yicha imkoniyatlari o'rganildi. Tibbiyot, huquq va ta'lim sohalaridagi NLP dasturlari yuqori aniqlikka ega bo'lsa-da, badiiy matn va ijodiy yozuvlar bilan ishlashda ba'zi cheklovlar mavjudligi kuzatildi. Bu esa kelajakdagi tadqiqotlar uchun yangi yo'nalishlar ochadi.

Xulosa. Ushbu tadqiqot natijalari shuni ko'rsatadiki, kompyuter lingvistikasi zamonaviy tilshunoslik va axborot texnologiyalari kesishmasida rivojlanayotgan fundamental yo'nalishlardan biridir. Tadqiqot davomida kompyuter lingvistikasi tamoyillari, uning asosiy usullari va zamonaviy texnologiyalar bilan uyg'unligi atroflicha o'rganildi. Olingan natijalarga asoslanib, bu sohaning nafaqat tilshunoslik tadqiqotlarida, balki real hayotda ham keng amaliy ahamiyatga ega ekanligi aniqlandi. Kompyuter lingvistikasi sohasida olib borilgan ilmiy izlanishlar shuni ko'rsatadiki, tabiiy tilni qayta ishlash texnologiyalari inson va mashina o'rtasidagi muloqotni soddalashtirish, tarjima tizimlarini rivojlantirish, avtomatlashtirilgan matn tahlili va sun'iy intellekt asosida ishlaydigan til modellarini yaratishda muhim rol o'ynaydi. Bugungi kunda neyron tarmoqlar, mashinaviy o'rganish algoritmlari va katta ma'lumotlar tahlili ushbu sohaning samaradorligini yanada oshirishga yordam bermoqda. Shuningdek, tadqiqot jarayonida kompyuter lingvistikasining matnni

автоматик qayta ishlashdagi roli o'rganilib, uning lingvistik tadqiqotlarga ta'siri chuqur tahlil qilindi. Ma'lumotlarga asoslanib, sohaning asosiy yo'nalishlari, jumladan, morfologik tahlil, sintaktik tuzilish, semantik modellar va pragmatik jihatlarining tadqiqi yetarlicha rivojlanib borayotganligi aniqlandi. Ushbu texnologiyalar tabiiy tilni chuqur tushunish va tahlil qilishga imkon yaratadi, natijada avtomatlashtirilgan tarjima, ovoz bilan boshqarish, matnni indeksatsiyalash va ma'lumotlarni ekstraksiya qilish kabi jarayonlar yanada takomillashadi.

Ushbu yondashuv ilm-fan va texnologiya rivojiga katta hissa qo'shib, zamonaviy jamiyatda axborot oqimlarini samarali boshqarish imkonini beradi. Kelajakda bu sohadagi tadqiqotlarni yanada chuqurlashtirish va yangi innovatsion yondashuvlarni ishlab chiqish zarur. Xususan, tabiiy tilni qayta ishlash algoritmlarining yanada aniqroq bo'lishi, til modellari va sun'iy intellekt texnologiyalarining inson nutqini tushunish darajasini oshirish muhim vazifalardan biri bo'lib qolmoqda. Shu sababli, kompyuter lingvistikasining yanada rivojlanishi ham nazariy, ham amaliy jihatdan muhim ilmiy yo'nalish bo'lib qolaveradi

Foydalanilgan adabiyotlar ro'yxati

1. Jurafsky, D., & Martin, J. H. (2021). Speech and Language Processing. Pearson.
2. Goldberg, Y. (2017). Neural Network Methods for Natural Language Processing. Morgan & Claypool Publishers.
3. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). "Attention Is All You Need". Advances in Neural Information Processing Systems (NeurIPS), 30.
4. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". Proceedings of NAACL-HLT 2019.
5. Hochreiter, S., & Schmidhuber, J. (1997). "Long Short-Term Memory". Neural Computation, 9(8), 1735-1780.
6. Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press.

7. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). "Distributed Representations of Words and Phrases and Their Compositionality". Advances in Neural Information Processing Systems (NeurIPS), 26.
8. Pennington, J., Socher, R., & Manning, C. D. (2014). "GloVe: Global Vectors for Word Representation". Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1532-1543.
9. Brown, P. F., Pietra, V. J. D., Pietra, S. A. D., & Mercer, R. L. (1993). "The Mathematics of Statistical Machine Translation: Parameter Estimation". Computational Linguistics, 19(2), 263-311.
10. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). "Language Models are Few-Shot Learners". OpenAI Research Paper.