



O'ZBEK TILIDAGI GENERATIV MATNLARNING SEMANTIK KO'RSATKICHLARI VA GALLYUTSINATIV MA'NO SURILISHLARI TAHLILI

To'ychiyev Abror Xabibulla o'g'li

Renessans ta'lim universiteti

Assistent o'qituvchisi

Tuychiyev13@gmail.com

+998932963378

Annotatsiya. Ushbu maqolada sun'iy intellekt va katta til modellari (LLM) tomonidan o'zbek tilida yaratiladigan generativ matnlarning semantik xususiyatlari, mantiqiy izchilligi hamda ularda kuzatiladigan gallyutsinativ ma'no surilishlari ilmiy va amaliy jihatdan tahlil qilinadi. Tadqiqotda o'zbek tilining agglutinatив tabiatidan kelib chiquvchi morpho-semantik muammolar va ularning sun'iy intellekt matn tuzilishiga ta'siri yoritilgan. Shuningdek, ma'no surilishi (semantic drift) va gallyutsinatsiyalarning turlari, sabablari va ularni bartaraf etish usullari ko'rib chiqiladi.

Kalit so'zlar: Generativ AI, katta til modellari (LLM), semantik ko'rsatkichlar, gallyutsinatsiya, ma'no surilishi, kompyuter lingvistikasi, o'zbek tili NLP.

АНАЛИЗ СЕМАНТИЧЕСКИХ ПОКАЗАТЕЛЕЙ И ГАЛЛЮЦИНАТОРНЫХ СДВИГОВ СМЫСЛА В ГЕНЕРАТИВНЫХ ТЕКСТАХ НА УЗБЕКСКОМ ЯЗЫКЕ

Аннотация. В данной статье научно и практически анализируются семантические особенности, логическая последовательность и наблюдаемые галлюцинаторные сдвиги смысла в генеративных текстах, создаваемых искусственным интеллектом и крупными языковыми





моделями (LLM) на узбекском языке. В исследовании освещаются морфосемантические проблемы, вытекающие из агглютинативной природы узбекского языка, и их влияние на структуру текста искусственного интеллекта. Также рассматриваются типы, причины и методы устранения смысловых сдвигов (semantic drift) и галлюцинаций.

Ключевые слова: Генеративный ИИ, большие языковые модели (LLM), семантические показатели, галлюцинация, смысловой сдвиг, компьютерная лингвистика, NLP узбекского языка.

ANALYSIS OF SEMANTIC INDICATORS AND HALLUCINATORY MEANING SHIFTS IN GENERATIVE TEXTS IN UZBEK

Abstract. This article scientifically and practically analyzes the semantic features, logical consistency, and hallucinatory meaning shifts observed in generative texts produced by artificial intelligence and large language models (LLMs) in the Uzbek language. The study highlights morpho-semantic problems arising from the agglutinative nature of the Uzbek language and their impact on AI text structure. Furthermore, types, causes, and mitigation strategies for semantic drift and hallucinations are examined.

Keywords: Generative AI, large language models (LLM), semantic metrics, hallucination, semantic drift, computational linguistics, Uzbek NLP.

Kirish. So'nggi yillarda katta til modellari (Large Language Models — LLM) va generativ sun'iy intellekt texnologiyalarining jadal rivojlanishi tabiiy tilga ishlov berish (NLP) sohasida tub burilish yasadi. Biroq, kam resursli (low-resource) va mantiqiy-grammatik qurilishi murakkab bo'lgan dillarda, xususan, o'zbek tilida generativ matnlarning aniqligi va semantik barqarorligi yetarlicha yuqori darajada emas. O'zbek tili agglutinativ xarakterga ega bo'lib, affiksatsiya



mexanizmi orqali bitta o'zakdan ko'plab yangi ma'no va sintaktik munosabatlar shakllanadi. Generativ modellar matn yaratishda ko'pincha statistik va ehtimollik qonuniyatlariga tayanadi, bu esa ba'zan mantiqiy va factual (haqiqatga to'g'ri keluvchi) xatoliklar — gallyutsinatsiyalar hamda semantik ma'no surilishlariga olib keladi. Ushbu tadqiqotning maqsadi — o'zbek tilida generatsiyalangan matnlarning semantik sifat ko'rsatkichlarini aniqlash va gallyutsinativ ma'no surilishlarining asosiy sabablarini tahlil qilishdan iborat.

Generativ matnlarning semantik ko'rsatkichlari. Generativ matn sifati faqat grammatik to'g'rilik bilan emas, balki chuqur semantik moslik bilan ham o'lchanadi. O'zbek tilidagi matnlarni tahlil qilishda quyidagi asosiy semantik ko'rsatkichlar muhim o'rin tutadi: Semantik kogerentlik va gegemonlik (Coherence & Cohesion): Matndagi gaplar va abzaslar o'rtasidagi mantiqiy bog'liqlik. O'zbek tilida ergashgan qo'shma gaplar hamda kelshik va egalik qo'shimchalarining to'g'ri zanjirlanishi matnning tutashligini ta'minlaydi.

Kontekstual moslik (Contextual Relevance): Prompt (so'rov)da berilgan konseptual g'oyaning matn davomida saqlanib qolishi. Lexico-semantic boylik va aniqlik: Bir xil so'zlarning takrorlanmasligi, konnotativ ma'nolarning to'g'ri tanlanishi hamda o'zbek tiliga xos bo'lgan ibora va frazeologizmlardan o'rinli foydalanilishi. Faktologik aniqlik (Factuality & Truthfulness): Generatsiyalangan ma'lumotlarning real obyektiv borliqqa va berilgan manbalarga mos kelishi.

Gallyutsinativ ma'no surilishlari. Generativ modellar matn ketma-ketligini bashorat qilishda "gallyutsinatsiya" hodisasini yuzaga keltirishi mumkin. O'zbek tili kontekstida gallyutsinatsiyalar va ma'no surilishlari asosan quyidagi turlarda namoyon bo'ladi:

1. Faktologik gallyutsinatsiya (Factual Hallucination)





Modellar o'zbek tili korpusida ma'lumot yetarli bo'lmaganda inventiv (uydirma) faktlar yaratadi. Masalan, mavjud bo'lmagan ilmiy asarlar, tarixiy sanalar yoki shaxslar haqida ishonchli uslubda noto'g'ri ma'lumot berish.

2. Lexico-semantic va morfologik ma'no surilishi

O'zbek tilidagi qo'shimchalar va so'z yasovchi morfemalarni noo'rin qo'llash natijasida so'zning tub ma'nosi o'zgarib ketadi. Misol: "Tadqiqotchi ushbu nazariyani isbotladi" o'rniga modellar kontekstga mos kelmaydigan "Tadqiqotchi ushbu nazariyani asoslantirilmadi" kabi inkor va faza ziddiyatlarini keltirib chiqarishi.

3. Lingvo-madaniy va kalka ma'no surilishlari (Translational Drift)

Modellarning aksariyati ingliz yoki rus tillaridagi ichki semantik strukturada (representation) fikrlab, uni o'zbek tiliga to'g'ridan-to'g'ri so'zma-so'z tarjima qiladi. Natijada o'zbek tili uslubiyatiga (stylistics) yot bo'lgan va ma'no mantiqsizligini keltirib chiqaruvchi konstruksiyalar hosil bo'ladi.

Gallyutsinasiya va ma'no surilishlarining sabablari

Sabab Faktorlari	Tavsifi va O'zbek Tiliga Ta'siri
Korpus hajmi va sifatining yetarsizligi (Data Scarcity)	O'zbek tilidagi raqamlashtirilgan, sifatli akademik va adabiy matnlar korpusining yetarlicha katta emasligi.
Tokenizatsiya muammolari	Agglutinativ tillarda standart tokenizatorlar (BPE, WordPiece) so'zlarni noto'g'ri qismlarga (subwords) bo'ladi. Bu esa semantik o'zak va grammatik ko'rsatkichlar o'rtasidagi aloqaning uzilishiga olib keladi.



Sabab Faktorlari	Tavsifi va O'zbek Tiliga Ta'siri
Kontekst darchasining yo'qolishi (Attention Bottleneck)	Uzoq matnlarni generatsiya qilish jarayonida model boshida berilgan konseptni unutib, ma'noni boshqa yo'nalishga burib yuborishi.

Muammoni bartaraf etish va semantik ko'rsatkichlari oshirish yo'llari. RAG (Retrieval-Augmented Generation) tizimlarini joriy etish: Matn generatsiya qilishdan oldin ishonchli o'zbekcha manbalar bazasidan (vektorli ma'lumotlar bazasi) ma'lumot qidirib topish va javobni faqat shu manbaga tayangan holda shakllantirish. Agglutinatив tillarga mos tokenizatsiyani rivojlantirish: O'zbek tili morfemik tahliliga asoslangan moliya va tilshunoslik uchun optimallashtirilgan tokenizatorlarni ishlab chiqish. RLHF (Reinforcement Learning from Human Feedback): O'zbek tili mutaxassislari (lingvistlar) tomonidan modellashtirilgan baholash va qayta aloqa tizimini yo'lga qo'yish.

Xulosa. O'zbek tilidagi generativ matnlarning semantik ko'rsatkichlarini oshirish va gallyutsinativ ma'no surilishlarini kamaytirish NLP sohasidagi dolzarb vazifalardan biridir. O'zbek tilining milliy-lingvistik xususiyatlarini va agglutinatив strukturasini hisobga oluvchi maxsus korpuslar va baholash metrikalarini (benchmark) yaratish generativ modellarning aniqligi va mantiqiy mustahkamligini ta'minlashda asosiy omil bo'lib xizmat qiladi.

FOYDALANILGAN ADABIYOTLAR

1. Vaswani, A., et al. (2017). Attention is All You Need. Advances in Neural Information Processing Systems.



2. Ji, Z., et al. (2023). Survey of Hallucination in Natural Language Generation. ACM Computing Surveys.
3. Abzalova, M. (2021). O'zbek tili kompyuter lingvistikasi va NLP asoslari. Toshkent.
4. Brown, T., et al. (2020). Language Models are Few-Shot Learners. NeurIPS.