

O'ZBEK TILIDAGI SO'ROVLARDA PRAGMATIK ROL MODELLASHTIRISH: KAM
RESURSLI LINGVISTIK MUHITDA SUN'IY INTELLEKT
GALLYUTSINATSIYALARINI KAMAYTIRISH STRATEGIYASI

Ifthor Lafasov

E-mail: yorqinalijonov484@gmail.com

<https://doi.org/10.5281/zenodo.21769069>

Annotatsiya. Maqola sun'iy intellekt gallyutsinatsiyalarini bartaraf etish muammosini kam resursli lingvistik muhit — o'zbek tili konteksti misolida tahlil qiladi. Tadqiqotning dolzarbligi shundaki, gallyutsinatsiyani kamaytirish bo'yicha mavjud usullarning aksariyati ingliz tilida ishlab chiqilgan va sinovdan o'tkazilgan, ularning boshqa tillarga o'tkazilishi esa avtomatik emas.

Maqolada ko'p tilli gallyutsinatsiya bo'yicha bir-biriga zid empirik ma'lumotlar tanqidiy tahlil qilinadi va bu ziddiyat metrikalar hamda abstensiya xulq-atvorining farqi bilan izohlanadi.

Amaliy natija sifatida beshta komponentdan iborat pragmatik rol protokoli taklif etiladi.

Protokolning cheklovlari va joriy etishdan oldin bajarilishi zarur bo'lgan validatsiya shartlari ochiq qayd etiladi.

Kalit so'zlar: kam resursli tillar, o'zbek tili, sun'iy intellekt gallyutsinatsiyasi, pragmatik rol, abstensiya, agglyutinativ morfologiya, tashqi bilim bilan boyitilgan generatsiya.

Abstract. The article analyzes the problem of mitigating artificial intelligence hallucinations within the context of a low-resource linguistic environment—specifically, the Uzbek language. The relevance of the study lies in the fact that the majority of existing hallucination mitigation methods have been developed and evaluated in English, making their transferability to other languages non-automatic. The paper critically evaluates contradictory empirical data regarding multilingual hallucinations, explaining these discrepancies through differences in evaluation metrics and abstention behavior. As a practical contribution, a five-component pragmatic role protocol is proposed. The limitations of the protocol and the validation prerequisites necessary prior to its implementation are explicitly outlined.

Keywords: low-resource languages, Uzbek language, AI hallucination, pragmatic role, abstention, agglutinative morphology, retrieval-augmented generation.

Аннотация. В статье анализируется проблема устранения галлюцинаций искусственного интеллекта на примере низкоресурсной лингвистической среды — контекста узбекского языка. Актуальность исследования заключается в том, что большинство существующих методов снижения галлюцинаций разработаны и протестированы на английском языке, а их переносу на другие языки не свойственен автоматический характер. В работе дается критический анализ противоречивых эмпирических данных по мультязычным галлюцинациям; данное противоречие объясняется различием метрик и поведения абстенции. В качестве практического результата предлагается пятикомпонентный протокол прагматической роли. В статье открыто отмечаются ограничения протокола и валидационные условия, необходимые для выполнения перед его внедрением.

Ключевые слова: *низкоресурсные языки, узбекский язык, галлюцинация искусственного интеллекта, прагматическая роль, абстенция, агглютинативная морфология, генерация, дополненная внешней информацией.*

1. KIRISH: Generativ sun'iy intellekt tizimlari O'zbekistonda ta'lim, tarjima, huquqiy maslahat va ma'muriy xizmatlar sohalarida tez sur'atda joriy etilmoqda. Bunday joriy etish gallyutsinatsiya xavfini alohida dolzarb qiladi: agar noto'g'ri, biroq ishonchli ohangda taqdim etilgan ma'lumot o'quv jarayoniga yoki huquqiy hujjatga kirs, uning oqibatlari bevosita moddiy va ijtimoiy zarar bilan o'lchanadi.

Muammoning o'ziga xosligi shundaki, gallyutsinatsiyani kamaytirishning tan olingan usullari — tashqi bilim bilan boyitilgan generatsiya, tekshiruv zanjiri, o'z-o'zini nazorat — asosan ingliz tilidagi ma'lumotlarda ishlab chiqilgan. Ularning o'zbek tiliga o'tkazilishi bir necha to'siqqa duch keladi: korpus hajmining cheklanganligi, agglyutinativ morfologiyaning tokenizatsiyaga ta'siri, ikki yozuv tizimining amalda birga qo'llanilishi va milliy domendagi tekshirish manbalarining raqamlashtirilmaganligi.

Tadqiqot maqsadi — o'zbek tilidagi so'rovlarda gallyutsinatsiyani kamaytirish uchun pragmatik rol modellashtirishga asoslangan amaliy protokolni ishlab chiqish va uning qo'llanilish chegaralarini belgilash.

2. MAVZUGA OID ADABIYOTLARNING TAHLILI

2.1. Ko'p tilli gallyutsinatsiya bo'yicha zid dalillar: Ushbu yo'nalishdagi adabiyot yagona xulosaga kelmagan va bu holatni ochiq qayd etish ilmiy halollik talabidir.

Birinchi pozitsiya: kam resursli tillarda faktuallik pastroq. Turli resurs darajasidagi tillarni qamragan tadqiqotlar erkin matn generatsiyasida kam resursli va yuqori resursli tillar o'rtasida sezilarli faktuallik tafovuti mavjudligini ko'rsatadi [1]. Mashina tarjimasida ham gallyutsinatsiya asosan kam resursli til juftliklarida kuzatilgan [1].

Ikkinchi pozitsiya: bunday bog'liqlik yo'q. Islam va hammualliflarning 30 ta til va oltita model oilasini qamragan yirik tadqiqotida gallyutsinatsiya darajasi bilan tilning resurs darajasi o'rtasida sezilarli korrelyatsiya aniqlanmagan [2].

Uchinchi pozitsiya: bog'liqlik teskari yo'nalishda ko'rinishi mumkin. NoMIRACL to'plamida ba'zi kam resursli tillar eng past gallyutsinatsiya ko'rsatkichini namoyish etgan; sabab esa modelning bu tillarda ko'proq "bilmayman" javobini berishi bilan izohlangan [3].

Muallifning tanqidiy tahlili. Uchinchi pozitsiya, aslida, birinchi ikkitasining chalkashligiga kalit beradi. Agar model kam resursli tilda tez-tez javob berishdan tiyilsa, gallyutsinatsiya ulushi pasayadi — ammo bu ishonchlilikning oshishi emas, **foydalilikning pasayishidir**. Demak, tadqiqotlar o'rtasidagi ziddiyat ko'p hollarda haqiqiy emas, metrik xarakterga ega: gallyutsinatsiyani abstensiyani hisobga olmasdan o'lchagan ishlar bilan uni hisobga olgan ishlar turli hodisani o'lchamoqda. Bundan amaliy xulosa kelib chiqadi: o'zbek tili uchun har qanday baholash **gallyutsinatsiya, abstensiya va foydalilik ko'rsatkichlarini bir vaqtda** qayd etishi shart.

2.2. Turkiy va morfologik murakkab tillar bo'yicha tajriba: Turk tili uchun maxsus gallyutsinatsiya aniqlash modellari ishlab chiqilgan bo'lib, mualliflar morfologik murakkab va kam resursli tillar uchun bunday vositalarning yetishmasligini alohida muammo sifatida ta'kidlaydilar [4]. Turkiy tillar oilasi doirasidagi moslashtirish masalalari bo'yicha nazariy ishlar o'zbek tilini ham qamrab olgan holda paydo bo'la boshladi [5]. Bu — ijobiy tendensiya, biroq o'zbek tili uchun tasdiqlangan gallyutsinatsiya benchmarklari hozircha mavjud emas; ushbu bo'shliq ilmiy jihatdan ochiq muammo sifatida qayd etiladi.

2.3. Abstensiya strategiyasi: Ko'p tilli sharoitda javobdan tiyilishga o'rgatish bo'yicha tadqiqotlar mavjud usullarning aksariyati faqat ingliz tilida baholanganini ko'rsatadi [6]. Bu — o'zbek tili uchun bevosita amaliy ahamiyatga ega bo'shliq.

3. TADQIQOT METODOLOGIYASI: Tadqiqot amaliy-tatbiqiy xarakterga ega. Qo'llanilgan usullar: adabiyotlarning tizimli tahlili; qiyosiy-lingvistik tahlil (o'zbek tilining tipologik xususiyatlarini promptlash talablari bilan bog'lash); protokol loyihalash (komponentli konstruksiya usuli); risk tahlili.

Tadqiqot obyekti — o'zbek tilidagi foydalanuvchi so'rovlariga generativ tizimlarning javob berish jarayoni. Predmeti — pragmatik rol konstruksiyalarining javob ishonchligiga ta'siri.

4. TAHLIL VA NATIJALAR

4.1. O'zbek tilining pragmatik rol loyihalashga ta'sir etuvchi xususiyatlari: Morfologik agglyutinatsiya. O'zbek tilida grammatik ma'no qo'shimchalar zanjiri orqali ifodalanadi. Bu tokenizatsiya bosqichida bir so'zning ko'p bo'lakka bo'linishiga olib keladi, natijada kontekst oynasi tezroq to'ladi va uzun manba matnlari bilan ishlash qiyinlashadi. Amaliy natija: o'zbek tilidagi promptlarda epistemik ko'rsatmalar qisqa va sintaktik sodda bo'lishi kerak — uzun ko'rsatma manba matni uchun ajratilgan joyni egallaydi.

Evidensiallikning grammatik ifodasi. O'zbek tilida ma'lumot manbasiga munosabat morfologik vositalar bilan ham ifodalanadi (-ibdi, ekan, emish kabi shakllar bilvosita, eshitilgan yoki tasdiqlanmagan ma'lumotni bildiradi). Bu — pragmatik rol loyihalash uchun **imkoniyat**: model javobida faktual va taxminiy mazmunni ajratish uchun sun'iy belgilarga emas, tilning o'z evidensial tizimiga tayanish mumkin. Ushbu imkoniyat, muallif bilishicha, hozirgacha empirik sinovdan o'tkazilmagan va tekshirishga muhtoj gipoteza maqomida qoladi.

Registr va ohang. O'zbek ilmiy va rasmiy uslubida qat'iy tasdiqlovchi ohang keng tarqalgan. Model bu uslubga taqlid qilganda, tasdiqlanmagan mazmun ham qat'iy shaklda beriladi — ya'ni uslubiy taqlid gallyutsinatsiya xavfini **kuchaytiradi**. Bu — identifikatsion rol ("siz professorsiz") ning salbiy ta'siri kuzatilishi mumkin bo'lgan aniq mexanizm.

4.2. PRAG-UZ protokoli: Yuqoridagi tahlil asosida beshta komponentdan iborat protokol taklif etiladi. Protokolning printsiplari shundaki, u modelga kim ekanini emas, qanday epistemik rejimda ishlashini ko'rsatadi.

K1. Manba chegarasi. Javob faqat taqdim etilgan hujjatga tayanishi belgilanadi. Tashqi bilimdan foydalanish taqiqlanadi yoki alohida belgilanishi talab qilinadi.

K2. Epistemik ajratish. Javobda uch toifa mustaqil belgilanadi: manbada bevosita tasdiqlangan; manbadan mantiqiy chiqarilgan; manbada mavjud bo'lmagan. O'zbek tilining evidensial shakllari bu ajratishning tabiiy vositasi sifatida qo'llaniladi.

K3. Abstensiya ruxsati. "Manbada bu haqda ma'lumot yo'q" javobi to'liq maqbul va afzal javob sifatida aniq belgilanadi. Bu komponent zarur, chunki ko'rsatmasiz model javob berish tomon og'adi.

K4. Verifikatsiya bosqichi. Javob shakllangandan so'ng har bir da'vo manbaga qarshi qayta tekshiriladi. Tekshiruvning samarali bo'lishi uchun tekshiruvchi bosqich dastlabki javobga emas, faqat da'vo va manbaga kirish huquqiga ega bo'lishi maqsadga muvofiq — aks holda tasdiqlash tomon og'ish yuzaga keladi [7].

K5. Til registri nazorati. Javob qat'iy tasdiqlovchi ohangdan tiyilishi va noaniqlik darajasini leksik hamda morfologik vositalar bilan ifodalashi talab qilinadi.

Muhim izoh. Protokol nazariy asosga ega, ammo o'zbek tilida empirik validatsiyadan o'tkazilmagan. Uni ishlab chiqarish muhitiga joriy etishdan oldin 4.4-bandda keltirilgan sinov dasturi bajarilishi shart.

4.3. Protokolning tashqi bilim bilan boyitilgan generatsiya bilan integratsiyasi: PRAG-UZ protokoli mustaqil yechim emas, balki RAG arxitekturasining pragmatik qatlami sifatida loyihalangan [8]. Mehnat taqsimoti quyidagicha: qidiruv moduli tegishli hujjatni topadi, protokol esa modelning topilgan hujjatga munosabatini tartibga soladi. Bu ajratish muhim, chunki RAG ning o'zi gallyutsinatsiyani bartaraf etmaydi — u faqat manba taqdim etadi; manbadan chetga chiqish ehtimoli saqlanib qoladi.

O'zbek tili uchun qo'shimcha cheklov: milliy domendagi ishonchli manbalarning (normativ-huquqiy hujjatlar, o'quv adabiyoti, statistik ma'lumotlar) raqamlashtirilgan va indekslashga tayyor hajmi cheklangan. Shu bois protokolning samaradorligi bevosita manba bazasining sifatiga bog'liq bo'lib, buni tashkiliy vazifa sifatida alohida hal qilish talab etiladi.

4.4. Sinov dasturi: Protokolni joriy etishdan oldin quyidagi sinov o'tkazilishi tavsiya etiladi:

1. O'zbek tilida kamida uch sohani (ta'lim, huquq, ma'muriy xizmat) qamragan, ekspertlar tomonidan tasdiqlangan savol-javob to'plami shakllantiriladi.

2. Uch shart taqqoslanadi: protokolsiz baza holati; faqat identifikatsion rolli holat; to'liq PRAG-UZ holati.

3. Uchta ko'rsatkich bir vaqtda qayd etiladi — gallyutsinatsiya ulushi, abstensiya ulushi, foydalilik ulushi.

4. Baholash kamida ikkita mustaqil ekspert tomonidan amalga oshiriladi; ekspertlar kelishuvi koeffitsiyenti hisoblanadi.

Kutilayotgan risk. Protokolning K3 komponenti abstensiyani rag'batlantirgani sababli, foydalilik ko'rsatkichi pasayishi mumkin. Agar sinov natijasida gallyutsinatsiyaning kamayishi foydalilikning shunga mutanosib pasayishi hisobiga yuz bergani aniqlansa, protokol samarali deb hisoblanmaydi va K3 komponenti qayta kalibrlanishi lozim. Bu shartni oldindan belgilash — natijani gipotezaga moslab izohlash xavfini kamaytiradi.

5. AMALIY TAVSIYALAR: Ta'lim muassasalari uchun. Generativ tizimlardan foydalanishda talabalar va o'qituvchilar uchun majburiy qoida: model javobidagi har bir raqam, sana, iqtibos va manba havolasi mustaqil tekshirilishi lozim. Ilmiy ish tayyorlashda modelga adabiyotlar ro'yxati tuzdirish amaliyoti taqiqlanishi tavsiya etiladi — bu soxta iqtiboslarning eng keng tarqalgan manbasidir.

Huquqiy va ma'muriy sohalar uchun. Generativ tizim faqat manbaga bog'langan rejimda, ya'ni normativ hujjatlar bazasiga ulangan holda qo'llanilishi kerak. Manbasiz erkin generatsiya rejimi bu sohalarida maqbul emas.

Tarjima faoliyati uchun. Kam resursli til juftliklarida gallyutsinatsiya xavfi yuqoriroq bo'lishi mumkinligi hisobga olinib, atamalar va tegishli ismlar bo'yicha majburiy inson nazorati saqlanishi lozim.

Ilmiy hamjamiyat uchun. O'zbek tili uchun ochiq gallyutsinatsiya benchmarkini yaratish ustuvor vazifa sifatida belgilanishi tavsiya etiladi. Bunday resurssiz har qanday baholash bilvosita va taxminiy bo'lib qolaveradi.

6. XULOSA: Kam resursli tillarda gallyutsinatsiya darajasi bo'yicha adabiyotdagi ziddiyat asosan metrik xarakterga ega: abstensiyaning hisobga olmagan o'lchovlar ishonchlilikning oshishini foydalilikning pasayishi bilan chalkashtiradi.

1. O'zbek tilining agglyutinativ tuzilishi kontekst resursiga qo'shimcha yuk beradi, evidensial morfologiyasi esa epistemik ajratish uchun tabiiy vosita taqdim etadi. Ikkinchi holat tekshiriladigan gipoteza maqomida qoladi.

2. O'zbek ilmiy-rasmiy uslubidagi qat'iy tasdiqlovchi ohang identifikatsion rol yuklashda gallyutsinatsiya xavfini kuchaytirishi mumkin — bu protokolda uslub nazorati komponentini zarur qiladi.

3. Taklif etilgan PRAG-UZ protokoli nazariy asosga ega bo'lsa-da, empirik validatsiyadan o'tkazilmagan; uning samaradorligi haqidagi har qanday da'vo sinovdan keyingina asosli bo'ladi.

Nazariy ahamiyat: pragmatik rol modellashtirish tipologik jihatdan farqli til muhitiga tatbiq etilishi konseptual jihatdan asoslandi.

Amaliy ahamiyat: ta'lim, huquq va tarjima sohalarida generativ tizimlarni joriy etish uchun aniq protokol va risk baholash mezonlari taklif etildi.

Istiqbolli yo'nalishlar: (a) o'zbek tili uchun ochiq gallyutsinatsiya benchmarkini yaratish; (b) evidensial morfologiyaning epistemik ajratishdagi samaradorligini eksperimental tekshirish; (c) protokolni boshqa turkiy tillarga kengaytirish va tillararo taqqoslash; (d) milliy manba bazasini indekslash arxitekturasini ishlab chiqish.

FOYDALANILGAN ADABIYOTLAR

1. Trustworthy Machine Learning via Memorization and the Granular Long-Tail: A Survey on Interactions, Tradeoffs, and Beyond. — arXiv:2503.07501, 2025. (kam resursli tillardagi faktuallik tafovuti bo'yicha umumlashtirilgan ma'lumotlar keltirilgan)
2. Islam S.O., Lauscher A., Glavaš G. How Much Do LLMs Hallucinate across Languages? On Multilingual Estimation of LLM Hallucination in the Wild. — arXiv:2502.12769, 2025.

3. Thakur N., Bonifacio L., Zhang X. et al. NoMIRACL: Knowing When You Don't Know for Robust Multilingual Retrieval-Augmented Generation. — arXiv:2312.11361.
4. Turk-LettuceDetect: A Hallucination Detection Model Suite for Turkish RAG Applications. — arXiv:2509.17671, 2025.
5. Ibrahimzade O., Tabasaransky K. Cross-Lingual Transfer and Parameter-Efficient Adaptation in the Turkic Language Family: A Theoretical Framework for Low-Resource Language Models. — arXiv:2604.06202, 2026.
6. Feng S., Shi W., Wang Y. et al. Teaching LLMs to Abstain across Languages via Multilingual Feedback. — arXiv:2406.15948, 2024.
7. MARCH: Multi-Agent Reinforced Self-Check for LLM Hallucination. — arXiv preprint, 2026.
8. Lewis P., Perez E., Piktus A. et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks // Advances in NeurIPS 33. — 2020.
9. Ji Z., Lee N., Frieske R. et al. Survey of Hallucination in Natural Language Generation // ACM Computing Surveys. — 2023. — Vol. 55, No. 12. — P. 1–38.
10. Huang L., Yu W., Ma W. et al. A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions // ACM Transactions on Information Systems. — 2025. — Vol. 43, No. 2. — P. 1–55.