

ANALYSIS OF THE HALLUCINATION PROBLEM IN LARGE LANGUAGE MODELS AND METHODS FOR MITIGATING IT

Adxamov Doniyorbek Abdulxodi o'g'li

Datatruck Foreign Enterprise LLC

doniyorbek.adhamov@gmail.com

Abstract

This thesis analyzes the types of hallucinations that occur in large language models, their underlying causes, methods for detecting them, and the main approaches to mitigating them. The methods are classified according to the data, retrieval, generation, and verification stages, and their applicability to low-resource languages such as Uzbek is evaluated.

Keywords

Artificial intelligence, large language models, hallucination, RAG, fact-checking, low-resource languages.

KATTA TIL MODELLARIDA GALLYUTSINATSIYA MUAMMOSI VA UNI KAMAYTIRISH USULLARI TAHLILI

Adxamov Doniyorbek Abdulxodi o'g'li

"Datatruck" MChJ Xorijiy Korxona

doniyorbek.adhamov@gmail.com

Annotatsiya. Tezisdagi katta til modellarida uchraydigan gallyutsinatsiya hodisasining turlari, kelib chiqish sabablari, uni aniqlash usullari va kamaytirishning asosiy yo'llari tahlil qilingan. Usullar ma'lumot, qidiruv, generatsiya va tekshirish bosqichlari bo'yicha tasniflangan hamda o'zbek tili kabi kam resursli tillar uchun ularning qo'llanish imkoniyatlari baholangan.

Kalit so'zlar: sun'iy intellekt, katta til modellari, gallyutsinatsiya, RAG, faktlarni tekshirish, kam resursli tillar.

Katta til modellari (LLM) bugungi kunda ta'lim, tibbiyot, huquq va davlat xizmatlari sohalarida keng qo'llanilmoqda. Ular matnni tushunish, umumlashtirish va savollarga javob berishda inson darajasiga yaqin natijalar ko'rsatmoqda. Biroq ularning muhim kamchiligi – gallyutsinatsiya, ya'ni grammatik jihatdan to'g'ri va ishonchli ko'rinadigan, lekin faktlarga mos kelmaydigan yoki manbada tasdiqlanmagan matn yaratishdir [1, 2]. Masalan, model mavjud bo'lmagan qonun moddasiga havola berishi yoki ilmiy maqolaning muallifini noto'g'ri ko'rsatishi mumkin. Bu muammo modellarni yuqori mas'uliyatli sohalarida qo'llashni cheklaydi.

Gallyutsinatsiya turlari

Ilmiy adabiyotda gallyutsinatsiya bir necha mezon bo'yicha tasniflanadi. Manbaga munosabati bo'yicha u ikki turga bo'linadi [1]. Manbaga zid (intrinsic) gallyutsinatsiyada javob berilgan matnga qarama-qarshi bo'ladi. Manbada tasdiqlanmagan (extrinsic) gallyutsinatsiyada esa javob manbada yo'q ma'lumotni qo'shadi, bu ma'lumotni na tasdiqlash, na rad etish mumkin.

2025-yilgi sharhda gallyutsinatsiya mazmuni bo'yicha ham ikki turga ajratilgan [3]. Bilimga oid gallyutsinatsiya noto'g'ri faktlar, sanalar, nomlar va o'ylab topilgan havolalarda namoyon bo'ladi. Mantiqqa oid gallyutsinatsiyada esa model to'g'ri asoslardan noto'g'ri xulosa chiqaradi, masalan, hisob-kitobda yoki bir necha bosqichli mulohazada xatoga yo'l qo'yadi. Bu tasnif amaliy ahamiyatga ega, chunki har bir tur uchun turli kamaytirish usullari samarali bo'ladi.

Kelib chiqish sabablari

Gallyutsinatsiyaning asosiy sabablari quyidagilar [2]:

- o'qitish ma'lumotlarida xato, qarama-qarshi yoki eskirgan bilimlarning mavjudligi;
- model bilimining ma'lum sana bilan cheklanganligi: o'qitishdan keyin qabul qilingan hujjatlar haqida model bilmaydi;

–modelning keyingi so'zni ehtimollik asosida tanlashi, ya'ni "bilmayman" deyish o'rniga eng ehtimolli davomni yaratishga moyilligi;

–kam uchraydigan faktlarning modelda yetarlicha mustahkamlanmaganligi.

Kam resursli tillarda bu muammo yanada kuchayadi, chunki model bu tillarda kamroq va sifati pastroq ma'lumot bilan o'qitilgan bo'ladi [5]. Natijada model o'zbek tilidagi savolga javob berishda ko'pincha boshqa tildagi bilimni noto'g'ri "tarjima" qiladi yoki mahalliy kontekstni (masalan, O'zbekiston qonunchiligini) boshqa davlatlar amaliyoti bilan aralashtirib yuboradi.

Gallyutsinatsiyani aniqlash va o'lchash

Muammoni kamaytirish uchun avvalo uni o'lchay olish kerak. FActScore usulida uzun javob alohida "atomar" faktlarga ajratiladi va har bir fakt ishonchli bilim manbai bilan tekshiriladi; natija tasdiqlangan faktlar ulushi sifatida ifodalanadi [6]. SelfCheckGPT usuli esa tashqi manbasiz ishlaydi: model bir savolga bir necha marta javob beradi, javoblar o'rtasidagi nomuvofiqlik gallyutsinatsiya belgisi sifatida qabul qilinadi [7]. RAG tizimlari uchun RAGAS freymvorkidagi sodiqlik (faithfulness) metrikasi javobdagi da'volarning berilgan kontekst bilan tasdiqlangan ulushini hisoblaydi [8].

Kamaytirish usullari

Adabiyotlar tahlili asosida usullarni to'rt guruhga ajratish mumkin (1-jadval).

1-jadval

Gallyutsinatsiyani kamaytirish usullarining tasnifi

Bosqich	Usullar	Qisqacha mazmuni
Ma'lumot va o'qitish	Ma'lumotlarni tozalash, faktlarga yo'naltirilgan qo'shimcha o'qitish	Model bilimining sifatini oshiradi, lekin katta resurs talab qiladi
Qidiruv (RAG)	Leksik, semantik va gibril qidiruv	Javobni tashqi bilim bazasidagi hujjatlarga tayantiradi [4]

Bosqich	Usullar	Qisqacha mazmuni
Generatsiya	Manbaga havola talab qiluvchi ko'rsatma, bosqichma-bosqich fikrlash, javobdan bosh tortish	Model faqat berilgan kontekst asosida javob beradi
Tekshirish	Javobni gap darajasida manba bilan solishtirish, o'z-o'zini tekshirish	Tasdiqlanmagan da'volarni aniqlaydi va olib tashlaydi

Qidiruv bilan boyitilgan generatsiya (RAG) bilimga oid gallyutsinatsiyani kamaytirishning eng amaliy usuli hisoblanadi, chunki u modelni qayta o'qitishni talab qilmaydi va bilim bazasini yangilash orqali eskirgan bilim muammosini hal qiladi [3, 4]. Shu bilan birga, RAG to'liq yechim emas: noto'g'ri topilgan yoki ishonchsiz hujjat ham gallyutsinatsiyaga sabab bo'lishi mumkin. Mantiqqa oid xatolarni kamaytirish uchun esa bosqichma-bosqich fikrlash va tekshirish usullari samaraliroq [3]. Shuning uchun amaliy tizimlarda bir necha usulni birgalikda qo'llash tavsiya etiladi.

O'zbek tili uchun ahamiyati. O'zbek tili uchun eng istiqbolli yo'nalish – RAG va tekshirish usullarini birgalikda qo'llash. Bu usullar katta hajmdagi o'zbekcha o'qitish ma'lumotlarini talab qilmaydi: modelning o'zini qayta o'qitish o'rniga, ishonchli o'zbek tilidagi hujjatlar bazasi yaratiladi. Biroq ularni samarali qo'llash uchun o'zbek tilining agglyutinatib tabiatini ("hujjat", "hujjatlarimizdagi" kabi shakllar), lotin va kirill yozuvlarining parallel qo'llanilishini hamda "o" va "g" harflaridagi apostrofning turlicha yozilishini hisobga oluvchi qidiruv tizimlari kerak. Aks holda tizim kerakli hujjatni topa olmaydi va model yana o'z "xotirasiga" tayanib javob beradi.

Xulosa. Gallyutsinatsiyani to'liq yo'q qilish hozircha mumkin emas, real maqsad uni kamaytirishdir [3]. O'zbek tilidagi amaliy tizimlar uchun qidiruv, ishonchli generatsiya va javobni tekshirish bosqichlarini birlashtirgan ko'p bosqichli yondashuv eng maqbul hisoblanadi. Bunday tizimlarni baholash uchun o'zbek tilidagi maxsus test to'plamlarini yaratish va FActScore, sodiqlik kabi metrikalarni o'zbek tiliga moslashtirish dolzarb vazifa bo'lib qolmoqda.

Foydalanilgan adabiyotlar

1. Ji Z., Lee N., Frieske R. va boshq. Survey of Hallucination in Natural Language Generation // ACM Computing Surveys. – 2023. – Vol. 55, No. 12. – P. 1–38.
2. Huang L., Yu W., Ma W. va boshq. A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions // ACM Transactions on Information Systems. – 2025. – Vol. 43, No. 2.
3. Li Y., Fu X., Verma G., Buitelaar P., Liu M. Mitigating Hallucination in Large Language Models (LLMs): An Application-Oriented Survey on RAG, Reasoning, and Agentic Systems // arXiv preprint arXiv:2510.24476. – 2025.
4. Lewis P., Perez E., Piktus A. va boshq. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks // Advances in Neural Information Processing Systems (NeurIPS). – 2020. – Vol. 33. – P. 9459–9474.
5. Hossain I., Banik M., Al Farid F., Uddin J., Karim H.A. Advancing Large Language Models for Low-Resource Languages: A Systematic Review of Pretraining, Adaptation, and Ethical Challenges // Computer Modeling in Engineering & Sciences. – 2026. – Vol. 148, No. 1.
6. Min S., Krishna K., Lyu X. va boshq. FActScore: Fine-grained Atomic Evaluation of Factual Precision in Long Form Text Generation // Proceedings of EMNLP. – 2023. – P. 12076–12100.
7. Manakul P., Liusie A., Gales M.J.F. SelfCheckGPT: Zero-Resource Black-Box Hallucination Detection for Generative Large Language Models // Proceedings of EMNLP. – 2023. – P. 9004–9017.
8. Es S., James J., Espinosa-Anke L., Schockaert S. RAGAS: Automated Evaluation of Retrieval Augmented Generation // Proceedings of EACL: System Demonstrations. – 2024. – P. 150–158.