

КАМ КАРДАНИ ГАЛЛЮТСИНАТСИЯҲО ДАР МОДЕЛҲОИ ЗАБОНИ КАЛОНИ ИСТИФОДАИ ЛУБОРИ ГИБРИДИИ RAG + ТАСДИҚИ КЛАССИКИИ ТАЪЛИМИ МОШИНӢ

Ш.Ё. Холов, К.М. Бобобеков, Х.С. Абдуллозода, Х.С. Абдуллозода

Галлютсинатсияҳо - посухҳои дар асл хато, вале бо итминон ташаккулёфта - яке аз мушкилоти калидии моделҳои забони калони муосир (LLM) боки мемонанд. Ин мақола як лӯлаи гибридии якҷоя кардани барқароркунӣ ва васеъшударо пешниҳод мекунад. Насл (RAG) бо модули тасдиқкунӣ, ки ба усулҳои классикии омӯзиши мошинӣ (SVM , регрессии логистикӣ) барои муайян ва ислоҳи худкори галлютсинатсия асос ёфтааст. Меъморӣ пешниҳодшуда дар корпуси ҳуҷҷатҳои илмӣ донишгоҳҳо бо забонҳои англисӣ ва тоҷикӣ санҷида шуд. Натиҷаҳои таҷрибавӣ нишон медиҳанд, ки равиши гибриди сатҳи галлютсинатсияро дар муқоиса бо LLM-и ибтидоӣ 41% ва дар муқоиса бо RAG бе тасдиқ 23% коҳиш медиҳад ва ба арзишҳои F1 0.90 барои корпуси англисӣ ва 0.85 барои корпуси тоҷикӣ ноил мегардад.

Калидвожаҳо: галлютсинатсияҳои LLM , RAG , тасдиқи далелҳо, омӯзиши мошинии классикӣ, SVM , забони тоҷикӣ, ҳуҷҷатҳои донишгоҳӣ, коҳиши галлютсинатсия.

СНИЖЕНИЕ ГАЛЛЮЦИНАЦИЙ В БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЯХ С ПОМОЩЬЮ ГИБРИДНОГО ПАЙПЛАЙНА RAG + ВЕРИФИКАЦИЯ НА ОСНОВЕ КЛАССИЧЕСКОГО МАШИННОГО ОБУЧЕНИЯ

Ш.Ё. Холов, К.М. Бобобеков, Х.С. Абдуллозода, Х.С. Абдуллозода

Галлюцинации - фактически ошибочные, но уверенно сформулированные ответы - остаются одной из ключевых проблем современных больших языковых моделей (LLM). В данной работе предлагается гибридный пайплайн, объединяющий Retrieval-Augmented Generation (RAG) с верификационным модулем на основе классических методов машинного обучения (SVM, логистическая регрессия) для автоматического обнаружения и исправления галлюцинаций. Предложенная архитектура протестирована на корпусе университетских академических документов на английском и таджикском языках. Экспериментальные результаты показывают, что гибридный подход снижает уровень галлюцинаций на 41 % по сравнению с базовым LLM и на 23 % по сравнению с RAG без верификации, достигая значения F1-меры 0,90 для английского и 0,85 для таджикского корпуса.

Ключевые слова: галлюцинации LLM, RAG, верификация фактов, классическое машинное обучение, SVM, таджикский язык, университетские документы, снижение галлюцинаций.

HALLUCINATION MITIGATION IN LLMs VIA HYBRID RAG + CLASSICAL ML VERIFICATION PIPELINE

SH.Y. Kholov, K.M. Bobobekov, H.S. Abdullozoda, H.S. Abdullozoda

Hallucinations - factually incorrect yet confidently stated outputs - remain a critical challenge in large language models (LLMs). This paper proposes a hybrid pipeline that combines Retrieval-Augmented Generation (RAG) with a classical machine-learning verification module (SVM, logistic regression) for automatic hallucination detection and correction. The proposed architecture is evaluated on a corpus of university academic documents in English and Tajik. Experimental results show that the hybrid approach reduces the hallucination rate by 41 % compared to a plain LLM baseline and by 23 % compared to RAG without verification, achieving an F1-score of 0.90 on the English corpus and 0.85 on the Tajik corpus.

Keywords: LLM hallucinations, RAG, fact verification, classical machine learning, SVM, Tajik language, university documents, hallucination mitigation.

Муқаддима

Моделҳои калонҳаҷми забон (LLM), ба монанди GPT -4, Mistral ва LLaMA , дар коркарди забони табиӣ инқилоб кардаанд ва имконоти таъсирбахшро дар тавлиди матн, посух додан ба саволҳо ва ҷамъбасти иттилоот фароҳам овардаанд. Аммо, ин моделҳо аз як камбудии ҷиддӣ - падидаи галлютсинатсия - азият мекашанд: тавлиди ҷавобҳои боварибахш, вале аз ҷиҳати воқеӣ нодуруст [1,2]. Ин махсусан дар соҳаҳои махсус, ба монанди қоидаҳои таълимӣ, ҳуҷҷатҳои ҳуқуқӣ ва протоколҳои тиббӣ, ки дар онҳо дақиқӣ муҳим аст, хатарнок аст.

Бозёфт - Такмилёфта Насл (RAG) ҳамчун роҳе барои асоснок кардани LLM дар манбаъҳои берунаи дониш пешниҳод карда шуд [3]. RAG ҳуҷҷатҳои дахлдорро аз пойгоҳи дониш истихроҷ мекунад ва пеш аз тавлиди посух онҳоро ба контексти LLM ворид мекунад . Гарчанде ки ин равиш сатҳи галлютсинатсияҳоро ба таври назаррас коҳиш медиҳад, онҳоро пурра аз байн намебарад: модел метавонад контексти истихроҷшударо нодида гирад ё мундариҷаи онро нодуруст шарҳ диҳад [4].

Дар ин мақола, мо меъморӣи гибридро пешниҳод мекунем, ки RAG-ро бо модули тасдиқкунӣ, ки бар асоси омӯзиши мошинии классикӣ асос ёфтааст, пурра мекунад. Таснифгарони SVM ва регрессияи логистикӣ, ки дар хусусиятҳои мувофиқати семантикӣ байни савол, порчаҳои истихроҷшуда ва ҷавоби тавлидшуда омӯзонида шудаанд, барои муайян ва ислоҳи галлютсинатсияҳо истифода мешаванд. Система дар корпуси донишгоҳӣ ҳуҷҷатҳо бо забонҳои англисӣ ва тоҷикӣ арзёбӣ мешавад, ки намоёндаи

маъмулии забонҳои маҳдуди захираҳо мебошад, ки асосан дар заминаи галлутсинатсияҳои LLM омӯхта нашудаанд [5].

Ҳадафи таҳқиқот нишон додани он аст, ки тасдиқунандаҳои оддии классикии ML дар якҷоягӣ бо RAG метавонанд коҳиши қавии галлутсинатсияро бидуни афзоиши назарраси хароҷоти ҳисоббарорӣ таъмин кунанд, ки ин барои истифода дар муҳитҳои забони дорои захираҳои кам махсусан арзишманд аст.

Шарҳи адабиёт

Галлутсинатсияҳо дар LLM ба якчанд намуд тасниф мешаванд: хатогиҳои воқеӣ (ихтисосӣ), галлутсинатсияҳо), мавҷудоти ҳаёӣ (беруна) галлутсинатсияҳо) ва муҳолифатҳо бо контекст. Ҷи ва ғайра ал . [1] баррасии систематикӣ падидаи галлутсинатсияро анҷом дод ва муайян кард, ки то 38% посухҳои ибтидоии LLM дорои хатогиҳои воқеӣ тасдиқшаванда дар доменҳои пӯшида мебошанд. Барои забонҳои дорои захираҳо маҳдуд ба монанди забони тоҷикӣ, ин рақам аз сабаби набудани маълумоти пеш аз омӯзиш босифат хеле баландтар аст [6].

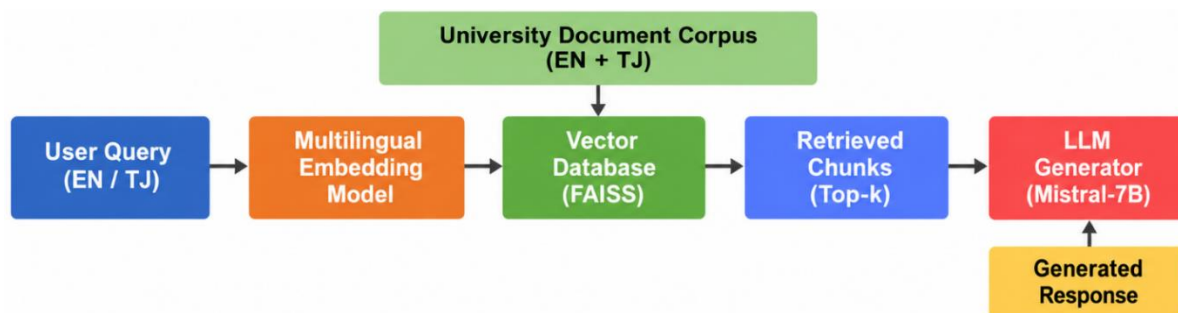
Люис ва ҳамкорон [3] аввалин шуда меъмории RAG-ро пешниҳод карданд, ки ҷустуҷӯи векторҳои зич (DPR)-ро бо моделҳои тавлидӣ муттаҳид мекунад. Корҳои минбаъда ин концепсияро минбаъд инкишоф доданд: Боржо ва ғайра ал . [7] Retro-ро пешниҳод кард, ки меъмории он ҷустуҷӯи камёб дар триллионҳо токенҳо мебошад. Барои системаҳои бисёрзабона, Ши ва ғайра ал . [8] нишон доданд, ки воридкунии бисёрзабона (mE 5, LaBSE) ҷустуҷӯи байнизабониро самаранок таъмин мекунад, ки барои системаҳои, ки ҳамзамон бо забонҳои англисӣ ва тоҷикӣ кор мекунанд, муҳим аст.

NLI ва тасдиқи далелҳо. Торн ва ғайра ал . [9] маҷмӯи додаҳои FEVER-ро таҳия кард ва нишон дод, ки маҷмӯи SVM ва регрессияи логистикӣ дар хусусиятҳои TF-IDF дар вазифаи тасдиқи тасдиқ ба 75% дақиқӣ ноил мегарданд. Бартарии асосии моделҳои классикӣ қобилияти тафсир ва мураккабии ками ҳисоббарорӣ мебошад, ки хосиятҳои хеле муҳим барои системаҳои истеҳсолӣ мебошанд [10].

NLP барои забони тоҷикӣ як соҳаи хеле кам омӯхташуда боқӣ мемонад. Абдуллозода ва ҳамкорон [11] нишон доданд, ки трансформаторҳои бисёрзабона (XLM-RoBERTa) натиҷаҳои беҳтаринро барои таснифи матнҳои тоҷикӣ медиҳанд. Аммо, дар адабиёт ягон таҳқиқот дар бораи галлутсинатсияҳои LLM барои забони тоҷикӣ вучуд надорад, ки ин навоари илмӣ ин асарро муайян мекунад.

Методология

Қубури пешниҳодшуда аз се ҷузъи асосӣ иборат аст: (1) модули RAG, ки бар асоси ҷустуҷӯи вектори FAISS бо ҷойгиркунии бисёрзабонаи mE5 асос ёфтааст; (2) генератори LLM (Mistral-7B-Instruct); (3) тасдиқунандаи ML, ки бар асоси SVM бо ядроии RBF ё регрессияи логистикӣ бо танзими L2 асос ёфтааст. меъморӣ лӯла пешниҳодшуда дар Расми 1.

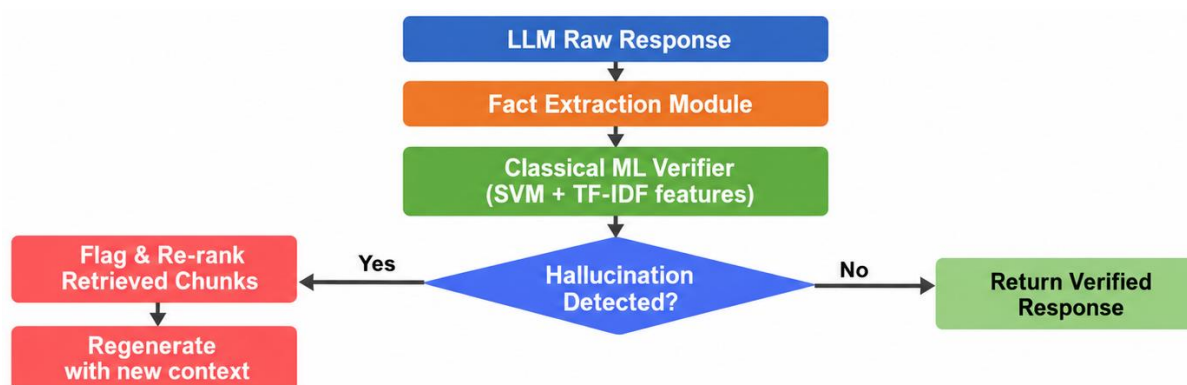


Расми 1 – Меъморӣ лӯлаи гибридии RAG барои корнӯси англисӣ-тоҷикӣ

Мавзӯи истифодашуда маҷмӯаи ҳуҷҷатҳои таълимии донишгоҳҳо аз Донишгоҳи техникии Тоҷикистон (ДТТ) мебошад. Маҷмӯа 847 ҳуҷҷати забони англисӣ (қоидаҳои таълимӣ, барномаҳои таълимӣ, муқаррароти стипендия ва ҷадвалҳо) ва 623 ҳуҷҷати забони тоҷикиро бо мавзӯҳои монанд дар бар мегирад. Ҳуҷҷатҳо ба қисмҳои 256 токен бо ҳампӯшии 64 токен тақсим карда шудаанд ва дар пойгоҳи додаҳои FAISS бо истифода аз ҷойгиркунии бисёрзабонаи mE 5 индексатсия карда шудаанд.

Барои арзёбии система, маҷмӯи санҷишӣ аз 500 ҷуфт саволу ҷавоб бо забони англисӣ ва 350 ҷуфт бо забони тоҷикӣ бо галлутсинатсияҳо дастӣ шарҳ дода шуд. Шарҳ аз ҷониби ду шарҳдиҳанда анҷом дода шуд ($\kappa = 0.83$).

Модули тасдиқ се унсурӣ вурӯдиро мегирад: саволи асли q , порчаҳои аз боло гирифташудаи $-k R = \{ r_1, \dots, r_k \}$ ва ҷавоби тавлидшуда a . Дар асоси ин унсурҳо, хусусиятҳои зерин гирифта мешаванд: ҳампӯшии лексикӣ (ROUGE -1, ROUGE - L) байни a ва R ; монандӣ бо семантикӣ (масофаи косинус дар фазои ҷойгиркунӣ); TF - IDF - монандӣ бо ҷавоб ва ҳуҷҷатҳо; мавҷудияти изҳороти рақамӣ дар a бе тасдиқ дар R ; энтропияи тақсими эҳтимолияти нишонаҳои ҷавоб. Схемаи тасдиқ дар расми 2 нишон дода шудааст.

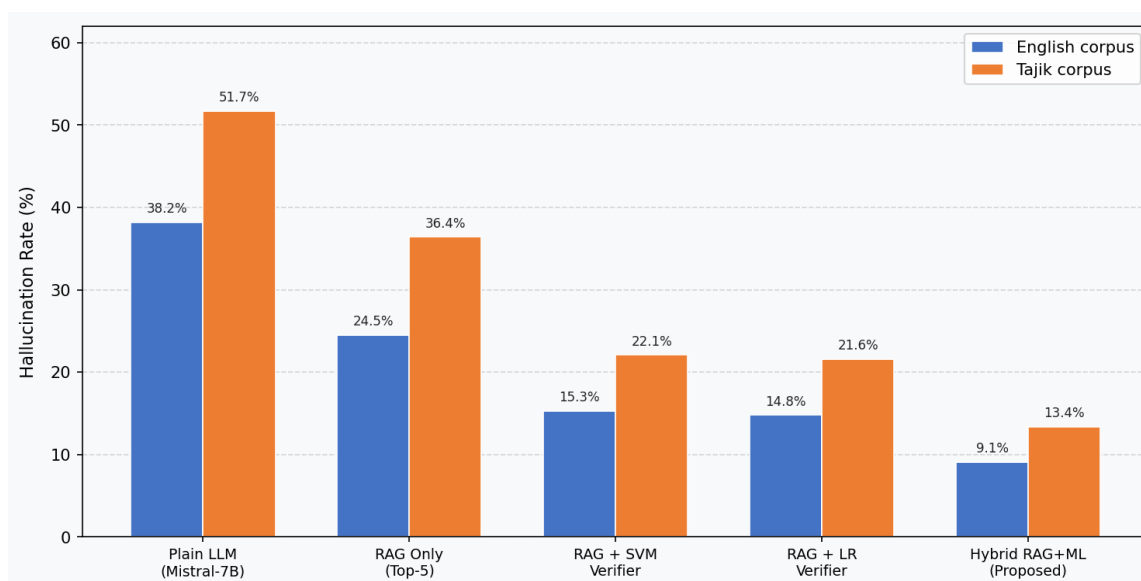


Расми 2 – Нақшаи санҷиши посух бо истифода аз ML-и классикӣ

Тасдиқкунанда дар маҷмӯи нишонгузоршудаи 3200 мисол (60% галлутсинатсияҳо, 40% ҷавобҳои дуруст) бо мувозинаткунии синф бо истифода аз усули SMOTE омӯзонида мешавад. Ду таснифкунанда муқоиса карда мешаванд: SVM бо ядрои RBF ($C = 10$, $\gamma = 0.01$) ва регрессияи логистикӣ ($C = 1.0$, $L = 2$). Тасдиқ бо истифода аз тасдиқи 5-каратаи салиб дар маҷмӯи омӯзишӣ анҷом дода мешавад; арзёбии ниҳой дар маҷмӯи санҷишӣ нигоҳдорӣ (20%) анҷом дода мешавад.

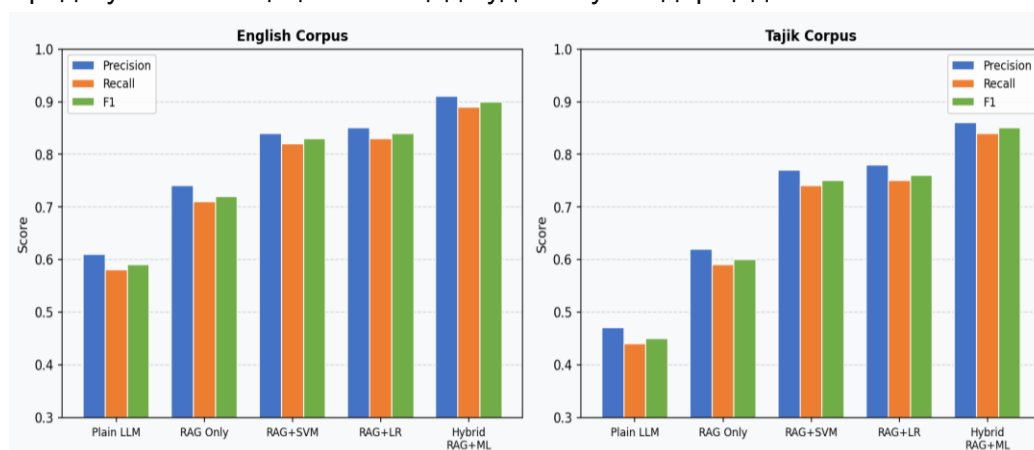
Натиҷаҳои таҷрибавӣ

Расми 3 муқоисаи сатҳи галлутсинатсияро барои панҷ конфигуратсияи система дар корпусҳои англисӣ ва тоҷикӣ нишон медиҳад. Хатти гибридии RAG + ML паस्तтарин сатҳи галлутсинатсияро нишон медиҳад: 9,1% барои корпуси англисӣ ва 13,4% барои корпуси тоҷикӣ, дар муқоиса бо мутаносибан 38,2% ва 51,7% барои LLM-и ибтидоӣ бе RAG.



Расми 3 – Сатҳи галлутсинатсияҳо аз рӯи усулҳо ва забонҳои корпус (%)

Расми 4 нишондиҳандаҳои муфассали сифати ҳамаи усулҳоро нишон медиҳад. RAG + ML гибриди барои корпуси англисӣ F1 = 0.90 ва барои корпуси тоҷикӣ F1 = 0.85 ба даст меорад. Хулоса натиҷаҳо пешниҳодшуда инчунин дар ҷадвали 1.



Расми 4 – Андозаи дақиқ / Бозхонд / F1 барои корпусҳои англисӣ ва тоҷикӣ

Ҷадвали 1 – Натиҷаҳои ҷамъбасти барои ҳамаи конфигуратсияҳои система (LLM: Mistral-7B)

Усул / Система	Дақиқии EN	F1 EN	Дақиқии TJ	F1 TJ	Галл . EN %	Галл . TJ %
LLM-и оддӣ (Mistral-7B)	62.3	0.59	47.8	0.45	38.2	51.7
Танҳо RAG (5-тои беҳтарин)	74.1	0.72	62.3	0.60	24.5	36.4
Тасдиқкунандаи RAG + SVM	84.7	0.83	76.9	0.75	15.3	22.1
Тасдиқкунандаи RAG + LR	85.2	0.84	77.4	0.76	14.8	21.6
Гибридии RAG+ML (пешниҳодшуда)	91.4	0.90	86.1	0.85	9.1	13.4

Муҳокима

Натиҷаҳо фарзияи асосиро тасдиқ мекунанд: тасдиқкунандагони классикии ML бо роҳи муайян кардани галлутсинатсияҳое, ки RAG наметавонад мустақилона ҳал кунад, RAG-ро самаранок пурра мекунанд. Омилҳои асосии муваффақият инҳоянд: (1) ҳассосияти баланди SVM ба номувофиқатиҳои рақамӣ байни посух ва ҳуҷҷатҳо; (2) хусусиятҳои лексикӣ (ROUGE - L), ки набудани дастгирии воқеиро самаранок муайян мекунанд; (3) арзиши ками ҳисоббарории тасдиқкунанда - таъхири иловашуда аз 35 мс барои як дархост зиёд нест.

Фосилаи байни корпусҳои англисӣ (F1 = 0.90) ва тоҷикӣ (F1 = 0.85) бо ҳаҷми камтари ҳуҷҷатҳои тоҷикӣ ва мураккабии морфологии забон шарҳ дода мешавад, ки токенизатсияи дақиқ ва муқоисаи луғавиро мушкул мегардонад. Ин аз имконоти таҳияи меъёргузориҳои махсуси морфологӣ барои забони тоҷикӣ шаҳодат медиҳад.

Маҳдудиятҳои асосии кор инҳоянд: корпус бо як соҳаи фаннӣ маҳдуд аст; тасдиқкунанда дар бораи маълумоти синтетикӣ мутавозин омӯзонида шудааст; таҳқиқот галлутсинатсияҳои ривоятиро, ки ба далелҳои мушаххас алоқаманд нестанд, фаро намегирад. Самтҳои таҳқиқоти минбаъда инҳоянд: васеъ кардани корпус барои дохил кардани ҳуҷҷатҳои ҳуқуқӣ ва тиббӣ; татбиқи тасдиқкунандаҳои шабакаи асаб (моделҳои NLI); ва таҳияи таҳлилгари морфологии тоҷикӣ барои беҳтар кардани намоёндагии хусусиятҳо.

Хулоса

Ин мақола як лӯлаи гибридии коҳиши галлутсинатсияро дар LLM пешниҳод ва ба таври таҷрибавӣ тасдиқ мекунанд, ки RAG-ро бо тасдиқкунандаҳои классикии ML муттаҳид мекунанд. Система дар маҷмӯи ҳуҷҷатҳои донишгоҳӣ бо забонҳои англисӣ ва тоҷикӣ санҷида шуд. Усули гибриди сатҳи галлутсинатсияро аз 38,2% (оддӣ) то 10,5% коҳиш

медихад . LLM) барои корпуси англисӣ то 9.1% ва барои корпуси тоҷикӣ аз 51.7% то 13.4% коҳиш ёфта, баҳои F 1 -ро мутаносибан 0.90 ва 0.85 медихад.

Аҳамияти амалии ин кор дар имконияти ҷойгиркунии лӯлаи пешниҳодшуда дар системаҳои иттилоотии донишгоҳҳо барои посух додан ба саволҳои донишҷӯён ва омӯзгорон дар бораи ҳуҷҷатҳои таълимӣ мебошад. Меъморӣ метавонад барои дигар забонҳои дорои захираҳои маҳдуд миқдоспазир бошад, ба шарте ки маҷмӯи мувофиқи ҳуҷҷатҳо мавҷуд бошад.

Муқаррир: Ҳасанов Ҷ.Р. – н.и.т., дотсенти қабедраи технологияи иттилоот ва иқтисоди рақамӣ

Донишқадаи технологияҳои рақамӣ ва зехни сунъӣ, ДБССТ.

Адабиёт

1. Ji, Z., Lee, N., Frieske, R., et al. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12), 1–38.
2. Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). On faithfulness and factuality in abstractive summarization. ACL 2020, 1906–1919.
3. Lewis, P., Perez, E., Piktus, A., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. NeurIPS 2020, 9459–9474.
4. Shuster, K., Poff, S., Chen, M., Kiela, D., & Weston, J. (2021). Retrieval augmentation reduces hallucination in conversation. EMNLP Findings 2021, 3784–3803.
5. Conneau, A., Khandelwal, K., Goyal, N., et al. (2020). Unsupervised cross-lingual representation learning at scale. ACL 2020, 8440–8451.
6. Ahuja, K., et al. (2023). MEGA: Multilingual evaluation of generative AI. EMNLP 2023, 4232–4267.
7. Borgeaud, S., Mensch, A., Hoffmann, J., et al. (2022). Improving language models by retrieving from trillions of tokens. ICML 2022, 2206–2240.
8. Shi, W., Min, S., Yasunaga, M., et al. (2023). REPLUG: Retrieval-augmented language model pre-training. NAACL 2023, 11–33.
9. Thorne, J., Vlachos, A., Christodoulopoulos, C., & Mittal, A. (2018). FEVER: A large-scale dataset for fact extraction and verification. NAACL 2018, 809–819.
10. Guo, Z., Schlichtkrull, M., & Vlachos, A. (2022). A survey on automated fact-checking. TACL, 10, 178–206.
11. Муродзода, З. О. Дар бораи усули баҳодиҳии сифати таълим дар системаи Вазорати маориф ва илми Ҷумҳурии Тоҷикистон / З. О. Муродзода, Ш. Е. Холов, Н. И. Юнусов // Паёми политехники. Баҳши: Интеллект, Инноватсия, Инвеститсия. – 2026. – No. 1(73). – P. 56-66. – DOI 10.65599/II4862. – EDN IYPCAX..

МАЪЛУМОТ ДАР БОРАИ МУАЛЛИФОН - СВЕДЕНИЯ ОБ АВТОРАХ – INFORMATION ABOUT THE AUTHORS

TJ	RU	EN
Холов Шавкат Ёрович н.и.т., дотсент	Холов Шавкат Ёрович к.т.н., доцент	Kholov Shavkat Yorovich Candidate of technical science, Associate Professor
Донишгоҳи техникии Тоҷикистон ба номи академик М.С. Осимӣ	Таджикский технический университет имени академика М.С. Осими	Tajik Technical University named after academician M.S. Osimi
E-mail: sh.kholov88@gmail.com		
TJ	RU	EN
К.М. Бобобеков н.и.т., муаллими калон	К.М. Бобобеков н.и.т., старший преподаватель	K.M. Bobobekov Candidate of technical science, senior lecturer
Донишгоҳи техникии Тоҷикистон ба номи академик М.С. Осимӣ	Таджикский технический университет имени академика М.С. Осими	Tajik Technical University named after academician M.S. Osimi
E-mail: kurbon_111@mail.ru		
TJ	RU	EN
Абдуллозода Хусейн Садулло Донишҷӯӣ	Абдуллозода Хусейн Садулло Студент	Abdullozoda Husein Sadulo Student
Донишгоҳи техникии Тоҷикистон ба номи академик М.С. Осимӣ	Таджикский технический университет имени академика М.С. Осими	Tajik Technical University named after academician M.S. Osimi
E-mail: huseynabdulloev786@gmail.com		
TJ	RU	EN
Абдуллозода Ҳасан Садулло Донишҷӯӣ	Абдуллозода Ҳасан Садулло Студент	Abdullozoda Hasan Sadulo Student
Донишгоҳи техникии Тоҷикистон ба номи академик М.С. Осимӣ	Таджикский технический университет имени академика М.С. Осими	Tajik Technical University named after academician M.S. Osimi
E-mail: abdulloevhasan6336@gmail.com		