

**LINGUISTIC MARKERS AND FAKE NEWS DETECTION:
INTEGRATING GLOBAL MODELS INTO UZBEK LANGUAGE
SPACE.**

Mo‘minova E‘zoza Muhammad qizi.

Ingliz filologiyasining 1-bosqich talabasi.

“Filologiya va xorijiy tillar kafedrası”,

Diplomat University, Toshkent.

Ilmiy rahbar: Mamadjanova Dildora Ulug‘bek qizi

ezomukhamedovna@gmail.com

Annotatsiya: Bugungi raqamli axborot asrida yolg‘on xabarlarining (fake news) shiddat bilan tarqalishi insonlarning ijtimoiy turmush tarziga va axborot xavfsizligiga jiddiy xavf solmoqda. Bunday yolg‘on xabarlar ko‘pincha ijtimoiy tarmoqlar orqali tarqatilib, keng auditoriyani o‘ziga jalb qilish va ularni manipulyatsiya qilish (ishontirish) maqsadini ko‘zlaydi. Raqamli texnologiyalar rivojlanib borayotgan bir paytda, tilshunos olimlar va kompyuter mutaxassislari matn tarkibidagi lingvistik markerlar orqali feyk xabarlarini avtomatik aniqlashning turli samarali usullarini ishlab chiqmoqdalar. Ushbu maqolada yolg‘on xabarlarining asosiy til belgilari hamda ularni matndan ajratib olish va aniqlashning zamonaviy usullari atroflicha tahlil qilinadi.

Kalit so‘zlar: Yolg‘on xabarlar, axborot xavfsizligi, lingvistik markerlar, sun‘iy intellekt, o‘zbek tili, intellektual filtrlar, raqamli texnologiyalar.

Abstract: In today's digital information age, the rapid dissemination of fake news poses a serious threat to people's social lifestyle and information security. Such fake news is often distributed through social networks, aiming to attract a wide audience and manipulate (persuade) them. At a time when digital technologies are advancing, linguists and computer specialists are developing various effective methods for the automatic detection of fake news through linguistic markers within the text. This article provides a comprehensive analysis of the primary linguistic features of fake news, as well as modern methods for extracting and detecting them from text.

Key words: Fake News, Information Security, Linguistic Markers, Artificial Intelligence, Uzbek Language, Intellectual Filters, Digital Technologies.

Аннотация: В сегодняшнюю эпоху цифровой информации стремительное распространение ложных новостей (fake news) представляет серьезную угрозу для социального образа жизни людей и информационной безопасности. Подобные ложные новости часто распространяются через социальные сети с целью привлечения широкой аудитории и манипулирования (убеждения) ею. В условиях

развития цифровых технологий лингвисты и компьютерные специалисты разрабатывают различные эффективные методы автоматического обнаружения фейковых новостей с помощью лингвистических маркеров в тексте. В данной статье всесторонне анализируются основные языковые признаки ложных новостей, а также современные методы их извлечения и обнаружения из текста.

Ключевые слова: Ложные новости, информационная безопасность, лингвистические маркеры, искусственный интеллект (ИИ), узбекский язык, интеллектуальные фильтры, цифровые технологии.

Kirish.

Bugungi raqamli axborot asrida ijtimoiy tarmoqlar, ayniqsa Telegram, Instagram, YouTube va boshqa platformalar hayotimizning ajralmas qismiga aylanib ulgurdi. Biroq ushbu platformalarda tarqalayotgan yolg'on xabarlar keng auditoriya orasida tez tarqalibgina qolmay, balki butun jamiyatning axborot xavfsizligiga ham jiddiy xavf solmoqda. Turli xil yolg'on ma'lumotlarning tarqalishi insonlar hayotiga va ruhiyatiga (psixologik jihatdan) salbiy ta'sir ko'rsatishi mumkin. Foydalanuvchilarning ma'lum bir qismi bunday xabarlarga qiziqib, ularni o'qishadi va mavzu ularga juda qiziqarli tuyuladi. Natijada ular xabarning rost yoki yolg'onligini tahlil qilmasdan, unga darhol ishonch bildiradilar, hatto ayrim hollarda ma'lumotning haqqoniyligini boshqa manbalardan tekshirib ham ko'rmaydilar. Yolg'on ma'lumotlarni tarqatishda foydalaniladigan bunday manipulyativ usullar bugungi kunda ochiq virtual makonda tez-tez uchrab turadi.

Shu sababli, ijtimoiy tarmoqlardagi yolg'on xabarlarni avtomatik ravishda aniqlash va tekshirishning samarali usullarini ishlab chiqish bugungi kunning dolzarb ilmiy vazifasi hisoblanadi. Zamonaviy texnologiyalar shiddat bilan rivojlanayotgan bir paytda, sun'iy intellekt yordamida feyk (yolg'on) xabarlarni avtomatik aniqlash muammosini lingvistik va dasturiy jihatdan o'rganagan yetakchi olimlardan biri K. Shu (Kai Shu, 2019) hisoblanadi [2]. Olim o'z tadqiqotlarida ijtimoiy tarmoqlardan olingan va omma orasida keng tarqalgan matnlarni sun'iy intellekt hamda mashinali o'rganish (Machine Learning) algoritmlari yordamida tahlil qilgan va ularning ishonchlilik darajasini aniqlash modellarini ishlab chiqqan. Shuningdek, V. Pérez-Rosas [1] (2018) va K. Shu [2] (2019) tomonidan olib borilgan ushbu fundamental tadqiqotlarga ko'ra, sun'iy intellekt yordamida matn tarkibidagi lingvistik markerlarni aniqlash yolg'on ma'lumotlarni fosh etishning eng samarali yo'nalishlaridan biri sifatida e'tirof etiladi. Olib borilgan izlanishlar natijasida olimlar ingliz tilidagi lingvistik ko'rsatkichlarga asoslangan maxsus qoidalar va NLP (tabiiy tilni qayta ishlash) modellarini muvaffaqiyatli ishlab chiqqanlar.

Xulosa qilib aytganda, xalqaro miqyosdagi ushbu zamonaviy yondashuvlarni o'zbek tili grammatikasi xususiyatlari bilan o'zaro moslashtirish va integratsiya qilish o'ta muhim

amaliy ahamiyatga ega. Mazkur tadqiqot natijalari kelgusida o'zbek tili internet makonida yolg'on ma'lumotlarni aniqlashga va filtrlashga xizmat qiluvchi milliy raqamli texnologiyalarni ishlab chiqish uchun mustahkam nazariy va amaliy asos bo'lib xizmat qilishi mumkin.

Shuningdek, ushbu ilmiy natijalar internet foydalanuvchilarini chalg'ituvchi kontentlardan himoya qila oladigan, real vaqt rejimida (real-time) faoliyat yurituvchi intellektual filtrlarni yaratishda ham muhim ahamiyat kasb etadi.

TADQIQOT METODOLOGIYASI VA EMPIRIK BAZA.

Mazkur tadqiqotda qiyosiy tahlil, kontent tahlili va lingvistik kuzatish metodlaridan foydalanildi. Xorijiy olimlar tomonidan yolg'on xabarlarni aniqlash bo'yicha ishlab chiqilgan lingvistik yondashuvlar o'rganilib, ular o'zbek tilining grammatik va uslubiy xususiyatlari bilan qiyoslandi. Tadqiqot davomida ijtimoiy tarmoqlarda uchraydigan axborot namunalariidagi lingvistik markerlar tahlil qilinib, ularning yolg'on ma'lumotlarni aniqlashdagi ahamiyati baholandi.

Tadqiqotning empirik bazasini shakllantirish maqsadida UzNet segmentidagi Telegram kanallari va Instagram sahifalaridan jami 300 ta matnli axborot namunasi yig'ib olindi. Obyektiv tahlilni ta'minlash uchun ushbu ma'lumotlar bazasi ikki qismga ajratildi:

135 ta verifikatsiyalangan (haqiqiy) matn: Rasmiy ommaviy axborot vositalari va davlat idoralari tomonidan tasdiqlangan neytral yangiliklar;

165 ta feyk (yolg'on) matn: Keyinchalik rasmiy ravishda rad etilgan, uydirma va manipulyativ xarakterga ega xabarlar.

Ingliz tilidan farqli o'laroq, o'zbek tilining agglyutinativ tabiati, ya'ni so'zlarga qo'shimchalar ketma-ket qo'shib borish xususiyati hamda ijtimoiy tarmoqlardagi o'ziga xos nutq uslubi yolg'on xabarlarning lingvistik markerlarini shakllantirishda alohida yondashuvni talab etadi.

K. Shuning 2019-yilda chop etilgan *"Beyond News Contents: The Role of Social Context for Fake News Detection"* nomli fundamental tadqiqotiga ko'ra, dezinformatsiyani muvaffaqiyatli identifikatsiya qilish uchun sun'iy intellekt tizimlariga matn tarkibidagi lingvistik ko'rsatkichlarni (linguistic markers) kiritish taklif etiladi [2]. Ushbu metodologik yondashuv asosida o'zbek tilidagi dezinformatsiya markerlari tizimi va ularni aniqlash mexanizmlari quyidagi jadvalda umumlashtirildi.

1-jadval. K.Shu modeli asosida o'zbek tilidagi dezinformatsiya markerlari klassifikatsiyasi.

Marker turi	Strukturaviy tavsifi va UzNet xususiyatlari	Identifikatsiya usuli (NLP algoritmlari)	Feyk matnlarda (n=165)	Haqiqiy matnlarda (n=135)
1.Eksplitsit (Ochiq) belgilar	Grafik anomaliyalar, CAPSLOCK (Katta harflar), ekspressiv punktuatsiya (!!!, ???), kalligrammalar ("shooook")	Grafik tahlil, tokenizatsiya va statistik hisoblash	75%	12%
2.Implitsit. (Yashirin) belgilar (Matnning ichki uslubi)''.	Affiksial ko'rsatkichlar (subyektivlik yuklovchi -da, -mi, -oq, -gina yuklamalari), "men/biz" olmoshlarining yashirilishi	Semantik tahlil (Machine Learning)	68%	12%
3.Klikbeyt elementlar.	Sun'iy emotsionallik yuklovchi va auditoriyani chalg'ituvchi leksik birliklar (" SHOK XABAR ", " TEZKOR ", " DAHSHAT ")	Leksik ma'lumotlar bazasi (Lexical Database) bilan solishtirish	80%	4%

1-jadvalda keltirilgan statistik tahlil natijalari shuni ko'rsatadiki, UzNet segmentidagi feyk xabarlarda klikbeyt elementlar (**80%**) va eksplitsit grafik anomaliyalar (**75%**) eng yuqori ko'rsatkichni qayd etgan. Haqiqiy (verifikatsiyalangan) yangiliklarda esa ushbu ko'rsatkichlar mos ravishda 4% va 12% ni tashkil etib, matnning neytral uslubda

yoziqlashini isbotlaydi. O'zbek tilining o'ziga xos xususiyati bo'lgan subyektivlik yuklovchi affiksial markerlar va shaxs olmoshlarining yashirilishi (implitsit belgilar) feyk matnlarning **68 foizida** aniqlandi. Bu esa o'zbek tilidagi dezinformatsiyani avtomatik aniqlashda nafaqat leksik bazaga, balki chuqur morfologik-sintaktik tahlilga ham tayanish zarurligini ko'rsatadi.

LINGVISTIK MARKERLARNING TAHLILI: Grafik va punktuatsion anomalialar. Ijtimoiy tarmoqlar tiliga xos bo'lgan kalligramma (so'zlarning vizual shakli bilan o'ynashish), alifbolar aralashuvi (lotin va kirill grafikalarining bir matnda qorishi) hamda ekspressiv punktuatsion anomalialar (!!!, ???) dezinformatsiyani identifikatsiya qiluvchi muhim determinant (aniqlovchi omil) bo'lib xizmat qiladi. Raqamli muhitda ushbu anomalialar oddiy imlo xatosi emas, balki manipulyativ ta'sir o'tkazish instrumenti hisoblanadi. Shuningdek, matn tarkibidagi klikbeyt elementlar va sun'iy emotsionallik yuklovchi leksik birliklar ("SHOK XABAR", "TEZKOR", "DAHSHAT") auditoriyani chalg'itishga qaratilgan ochiq eksplitsit markerlardir.

Implitsit markerlar va stilistik tahlil. Matnlarning chuqur stilistik va tonallik tahlili orqali tashqi tomondan neytral ko'ringan axborot qatlami ortiga berkingan yashirin manipulyativ-hissiy tonallikni ajratib olish modelning yakuniy aniqlik koeffitsiyentini (accuracy) sezilarli darajada oshiradi. Xususan, o'zbek tilidagi affiksial ko'rsatkichlar (subyektivlik yuklovchi *-da, -mi, -oq, -gina* kabi qo'shimchalar) va shaxs olmoshlarining (*men/biz*) ongli ravishda yashirilishi — matnning ichki uslubini manipulyatsiya qilishning asosiy usulidir.

Dezinformatsion matnlarda axborot odatda umumiy yoki qisqa gap ko'rinishida berilib, harakat bajaruvchi shaxslar aniq ko'rsatilmaydi. Buning asosiy sababi — soxta xabar muallifining tarqatilgan yolg'on axborot uchun mas'uliyatni o'z bo'yniga olmaslik va subyektini yashirish istagidir. Shu sababli, feyk xabarlarda subyektivlikni ifodalovchi "men" yoki "biz" olmoshlari juda kam ishlatiladi. Ularning o'rniga foydalanuvchilar diqqatini tortish va manipulyatsiya qilish maqsadida subyektsiz, shok xarakteridagi sarlavhalardan foydalaniladi. Telegram ijtimoiy tarmog'idagi ochiq manbalardan olingan materiallar tahlili shuni ko'rsatadiki, *"Shok xabar! U bugun tunda vafot etdi"* yoki *"U bugun noma'lum shaxs tomonidan o'ldirib ketilgani aytilmoqda"* kabi sarlavhalar bunga yaqqol namuna bo'la oladi. Bunday ekspressiv konstruksiyalar auditoriyani emotsional jihatdan o'ziga jalb qilib, qiziqish uyg'otadi.

Tadqiqotlar shuni ko'rsatadiki, olim K. Shu faqatgina matnning lingvistik xususiyatini yoki sof lingvistik kontentini tahlil qilish yetarli emas, degan xulosaga kelgan [2]. Ijtimoiy tarmoqlarda tarqalayotgan xabarlarga nisbatan foydalanuvchilar tomonidan bildirilayotgan shubha va salbiy fikrlar (ijtimoiy kontekst) integratsiyasi tekshirilganda, axborotning uydirma ekanligini aniqlash koeffitsiyenti yoki identifikatsiya aniqligi darajasi yuqori

bo'lishi isbotlandi. Ushbu fikrlarni umumlashtirgan holda shuni aytish mumkinki, yolg'on xabarlarini aniqlashda matn ichidagi lingvistik belgilar — olmoshlar va emotsiyalarga tayanganlik darajasini o'rganish hamda ularni sun'iy intellekt yordamida avtomatik tahlil qilish hozirgi kunda eng samarali yechimdir.

Xulosa.

Ijtimoiy tarmoqlar orqali tarqalayotgan yolg'on xabarlar nafaqat ommaning e'tiborini o'ziga jalb qiladi, balki raqamli muhitda jamiyatning axborot xavfsizligiga ham jiddiy tahdid soladi. Bunday dezinformatsion materiallarning keng tarqalishi foydalanuvchilarning ruhiy holatiga salbiy ta'sir ko'rsatishi mumkin. Ko'pincha internet foydalanuvchilari virtual makondagi xabarlarining haqqoniyligini chuqur tahlil qilmasdan, ko'proq hissiyotga berilgan holda ularni tezda ulashishga va tarqatishga moyil bo'ladilar. Shu sababli, UzNet segmentida dezinformatsiyani aniqlash va tekshirishning zamonaviy texnologik usullarini ishlab chiqish bugungi kunning dolzarb vazifasidir.

V. Pérez-Rosas (2018) va K. Shu (2019) kabi xorijiy olimlarning fundamental tadqiqotlariga tayangan holda, sun'iy intellekt yordamida lingvistik markerlarni aniqlash yolg'on xabarlarini fosh etishning eng samarali yo'nalishi ekanligi ushbu ishda ham o'z isbotini topdi. Mazkur tadqiqot doirasida xalqaro ilmiy qonuniyatlarni o'zbek tilining agglutinativ tabiati va o'ziga xos grammatik xususiyatlari bilan o'zaro sintez qilish muhim amaliy natijalarni berdi. Tadqiqot davomida shakllantirilgan xulosalar kelgusida o'zbek tili internet makonini (UzNet) soxta ma'lumotlardan real vaqt rejimida himoya qiluvchi milliy intellektual filtrlarni ishlab chiqish uchun muhim nazariy va amaliy poydevor bo'lib xizmat qiladi.

ADABIYOTLAR RO'YXATI

[1] V. Pérez-Rosas, B. Kleinberg, A. Lifshitz, and R. Mihalcea, "Automatic detection of fake news," in *Proceedings of the 27th International Conference on Computational Linguistics*, 2018, pp. 42–53.

[2] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Beyond news content: The role of social context for fake news detection," in *WSDM '19: Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, 2019, pp. 110–122.