O‘ZBEK TILIDAGI FIRIBGARLIK XABARLARINI ANIQLASHDA GIBRID YONDASHUV: QOIDALAR, PATTERNLAR VA ONNX NEYRON TARMOG‘INING BIRLASHUVI

Axmatov Bekzod Nurali o‘g‘li
Abdullayev Xushnud Raxmatulla o‘g‘li
Saidov O‘lmasbek Azamatovich
Ramazonova Marjona Ikrom qizi
Muhammad al-Xorazmiy nomidagi TATU talabalari

Annotatsiya: *Ushbu maqolada o‘zbek tilidagi SMS va messenjer xabarlari ichidan firibgarlik (phishing, smishing, scam) kontentini avtomatik aniqlash uchun gibrid yondashuv ko‘rib chiqilgan. Tizim uchta mustaqil tahlil qatlamini birlashtiradi: 32 ta qoidaga asoslangan signal agregatsiyasi, 20 dan ortiq oldindan tuzilgan signal kombinatsiyalarini aniqlovchi pattern dvigateli (pattern engine) va 15 003 o‘lchovli TF-IDF asosidagi xususiyat vektori bilan ishlaydigan ONNX formatidagi neyron tarmoq modeli. Yakuniy xavf bahosi RiskFusion algoritmi orqali shakllantiriladi, u uchta manbani — Pattern Floor, Signal Risk va ML skorni — ustuvorlik bo‘yicha birlashtiradi. Yondashuvning o‘ziga xos jihati shundaki, butun tahlil mobil qurilmaning o‘zida (on-device) bajariladi va xabar matni hech qachon serverga uzatilmaydi. Bu o‘zbek tilidagi kontentni mahalliy modellar bilan tahlil qilish, kirill-lotin homoglyph hujumlariga qarshi himoya va past kechikishli (real-time) ishlash imkonini beradi.*

Kalit so‘zlar: *firibgarlik aniqlash, fishing, smishing, o‘zbek tili, gibrid AI, qoidaga asoslangan tizim, pattern engine, neyron tarmoq, ONNX, TF-IDF, homoglyph hujumi, ijtimoiy muhandislik, RiskFusion, on-device AI, mobil xavfsizlik, matn klassifikatsiyasi.*

O‘zbek tilida tarqaladigan firibgarlik xabarlari bugungi kunda kiber-tahdidlarning eng katta ulushini tashkil etadi. Aksariyat hujumlar uchta klassik shaklda yetib keladi: smishing (SMS orqali fishing havola yuborish), brand spoofing (rasmiy bank yoki to‘lov tizimi nomidan soxta xabar) va social engineering (shoshilinch, qo‘rqitish yoki yutuq va’dasi orqali foydalanuvchini havolaga bosishga majbur qilish).

Globalda keng tarqalgan ingliz tiliga moslashtirilgan tijoriy yechimlar — Google Messages spam filtri, mobil operatorlarning anti-spam xizmatlari — mahalliy tildagi xabarlarni samarali tahlil qila olmaydi, chunki ular ingliz korpusiga o‘qitilgan va o‘zbek leksikasini, fleksiyasini hamda o‘zbek-rus aralash kontekstlarni tushunmaydi.

Bundan tashqari, mahalliy hujumlar o‘ziga xos texnik xususiyatga ega — homoglyph (gomoglif) hujumlari. Firibgarlar matnli filtrlardan qochish uchun lotin va kirill harflarining ko‘rinishi o‘xshash juftlaridan foydalanadi.

Masalan, lotin “o” (Unicode U+006F) o‘rniga kirill “о” (U+043E) yoziladi va so‘z foydalanuvchi ko‘zi uchun bir xil ko‘rinadi, lekin matnli model uchun butunlay boshqa belgi hisoblanadi.

“Click.uz” so‘zi “Click.uz” deb yozilganida (uchta harfi kirill), klassik string-matching qoidasi ishlamaydi.

Shu sababli o‘zbek tili uchun firibgarlik aniqlash tizimi quyidagi uchta sifatga bir vaqtda ega bo‘lishi shart: mahalliy leksikani tushunish, homoglyphlarga immun bo‘lish va past kechikish bilan (mobil qurilmada) ishlash.

Matn normalizatsiyasi — barcha keyingi qatlamlar uchun zarur asos. Tahlilning birinchi va ko‘pincha kamtarin ko‘rinadigan, lekin amalda hal qiluvchi bosqichi — matnni kanonik shaklga keltirish.

Bu bosqich hamma matnni kichik harfga o‘tkazadi, kirill harflarini lotin ekvivalentlariga avtomatik almashtiradi (homoglyph swap), turli xil apostrof belgilarini (’, ’, ’) bir xil belgiga keltiradi, URL-larni matn ichidan ajratib oladi va xavfli kengaytmalarni (.apk, .exe, .scr, .bat) maxsus token sifatida belgilaydi. Normalizatsiyadan o‘tgan matn keyingi qatlamlar uchun “toza”, taxmin qilinadigan kirish shaklini taqdim etadi.

Aynan shu bosqich tufayli homoglyph hujumlari samarasiz bo‘lib qoladi: kirill “а” lotin “a” ga o‘tkaziladi, “Click.uz” so‘zi “click.uz” ga aylanadi va keyingi pattern aniqlash qoidalari xabarni odatdagidek tahlil qiladi.

Niyat (intent) aniqlash — kontekstni belgilash.

Tozalangan matn niyat detektoriga uzatiladi. Niyat to‘rt sinfdan biriga taalluqli bo‘ladi: REQUEST (so‘rov: “shu raqamga 50 ming yuboring”), WARNING (ogohlantirish: “hisobingiz bloklandi”), INFO (oddiy axborot xabari: “100 000 so‘m o‘tkazildi”) yoki UNKNOWN.

Niyat o‘z-o‘zidan xavf darajasini belgilamaydi, lekin keyingi qatlamlar uchun muhim kontekst beradi.

Masalan, WARNING+REQUEST kombinatsiyasi — ya’ni “hisobingiz bloklandi, mana shu havoladan kirib parolni tasdiqlang” — klassik fishing namunasi va bu kombinatsiya pattern dvigatelida maxsus belgilanadi.

Aksincha, INFO niyati va bank rasmiy formatidagi ma’lumotlar safe-counter signaliga aylanadi va yakuniy xavf balansini pasaytiradi.

Qoidalarga asoslangan signal agregatsiyasi — eng aniq va tushuntiriladigan qatlam. Bu bosqich gibrid yondashuvning eng shaffof qismi.

Normalizatsiyadan o‘tgan matn ustida 32 ta mustaqil signal hisoblanadi, va har bir signal oltita kategoriyadan biriga tegishli bo‘ladi: MALWARE (xavfli kengaytmalar, APK havolasi, “o‘rnating” buyrug‘i), PHISHING (“hisobingiz bloklandi”, “parolni tasdiqlang”, soxta brand nomlari), FINANCIAL (pul ko‘chirish so‘rovi, karta raqami so‘rovi, OTP-kod so‘rovi), LINK (URL’lar soni, shortener’lar, shubhali yuqori-darajali domenlar — .xyz, .top, .click), SOCIAL_ENGINEERING (“shoshilinch”, “24 soat ichida”, “yutuq”, emotsional bosim) va INTENT (oldingi niyat detektoridan kelgan qiymat).

1.1-rasm. Qoidalarga asoslangan signal agregatsiyasi

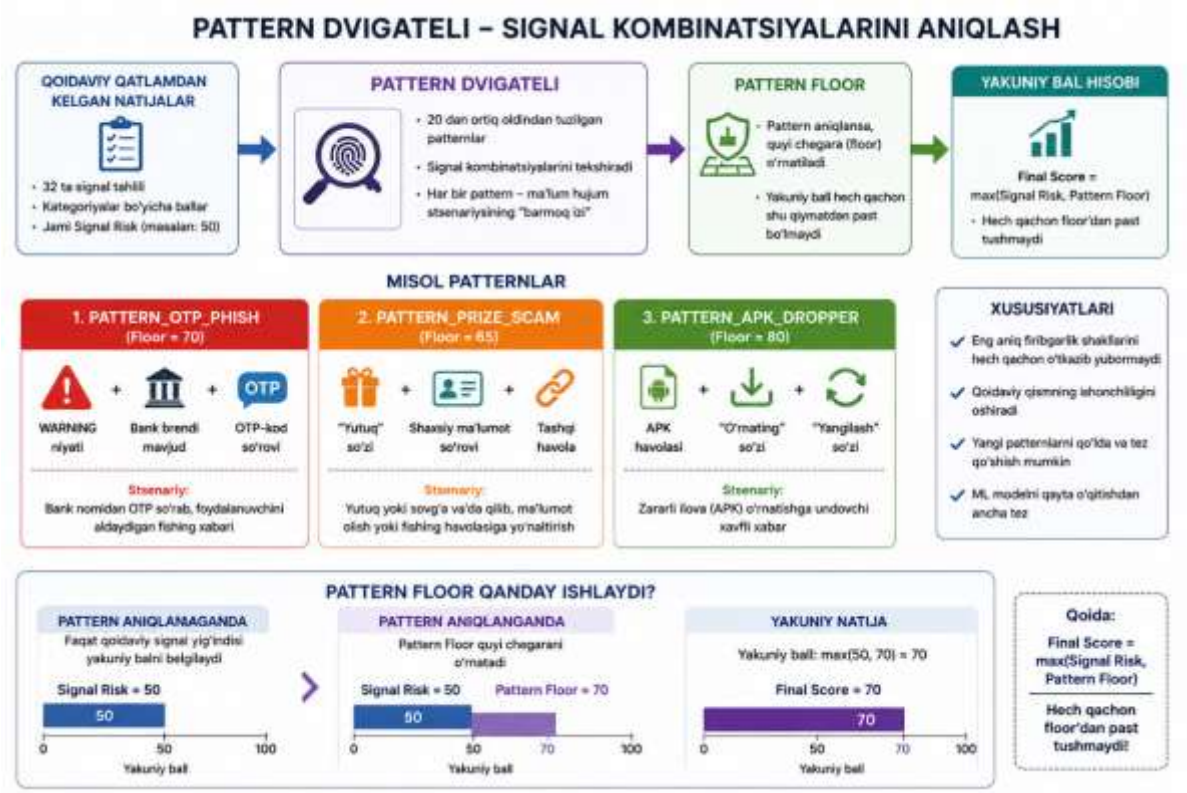
Har bir signal ma’lum bir og‘irlik qiymatiga ega. Qoidaviy mantiq oddiy: agar signal aniqlangan bo‘lsa, uning og‘irligi tegishli kategoriya hisobiga qo‘shiladi. Lekin har bir kategoriya o‘z maksimal chegarasiga (cap) ega. Masalan, SOCIAL_ENGINEERING kategoriyasi yakuniy balansga eng ko‘pi bilan 30 ball, FINANCIAL esa 40 ball qo‘sha oladi. Bu chegaralar maxsus o‘rnatilgan: ular bitta kategoriyaning yolg‘iz o‘zi xabarni “xavfli” deb e’lon qila olmasligini va xulosa kamida ikkita mustaqil sabab asosida shakllanishini majbur qiladi. Bundan tashqari, SAFE deb nomlangan qarama-qarshi kategoriya ham mavjud — agar matnda bankning rasmiy formatini ko‘rsatuvchi belgilar bo‘lsa (transaksiya raqami, sana, balans formati, rasmiy SMS-jo‘natuvchi nomi), SAFE balli o‘sadi va yakuniy xavf balansini pasaytiradi.

Aniq misol — qoidaviy qatlamning ishlashi. Quyidagi xabar: “Click.uz ogohlantiradi: 24 soat ichida tasdiqlamasangiz, kartangiz bloklanadi. Tasdiqlash: hxxp://click-uz.xyz/login” — quyidagi signallarni qo‘zg‘aydi: WARNING niyati (+8), “Click.uz” brand spoofing (+15, PHISHING), “24 soat ichida” — vaqt bosimi (+10, SOCIAL_ENGINEERING), “kartangiz bloklanadi” — moliyaviy qo‘rqitish (+12, FINANCIAL), “hxxp://” obfuskatsiyalangan havola (+8, LINK), “.xyz” shubhali TLD (+5, LINK). Cap’lar bilan jami signal yig‘indisi taxminan 58 ball, ya’ni SHUBHALI darajasiga yaqin. Lekin keyingi qatlamlar — pattern dvigateli va neyron tarmoq — bu xabarni aniq XAVFLI deb tasniflashga olib keladi.

Pattern dvigateli — signal kombinatsiyalarini aniqlash. Yakka signallar tahlilining bitta zaifligi bor: agar firibgar bittagina kuchli signaldan qochsa, jami balans pasayib ketadi va qoidaviy tizim xabarni o‘tkazib yuboradi. Pattern dvigateli ana shu zaiflikni yopadi. U 20 dan ortiq oldindan tuzilgan signal kombinatsiyalarini tekshiradi, va har bir pattern ma’lum bir hujum stsenariysining “barmoq izi” sifatida xizmat qiladi. Misol

uchun: PATTERN_OTP_PHISH — WARNING niyati bilan birga bank brendi va OTP so‘rovi mavjudligi, floor 70 ga teng; PATTERN_PRIZE_SCAM — “yutuq” so‘zi va shaxsiy ma’lumot so‘rovi va tashqi havolaning birgalikda paydo bo‘lishi, floor 65; PATTERN_APK_DROPPER — APK havolasi va “o‘rnating” va “yangilash” so‘zlarining ketma-ketligi, floor 80.

Pattern aniqlansa, u Pattern Floor o‘rnatadi — yakuniy balansning quyi chegarasi. Hatto agar boshqa signallar mustaqil ravishda 50 ball bersa-da, pattern floor 70 bo‘lsa, yakuniy bal max(50, 70) = 70 bo‘ladi. Bu mexanizm qoidaviy qismning eng ishonchli tomonidir, chunki u eng aniq belgilangan firibgarlik shakllarini hech qachon o‘tkazib yubormaydi. Pattern’lar tahlilchi muhandis tomonidan qo‘lda tuziladi va yangi hujum to‘lqinlari paydo bo‘lganda qisqa muddatda qo‘shilishi mumkin — bu ML modelni qayta o‘qitishdan ancha tez.



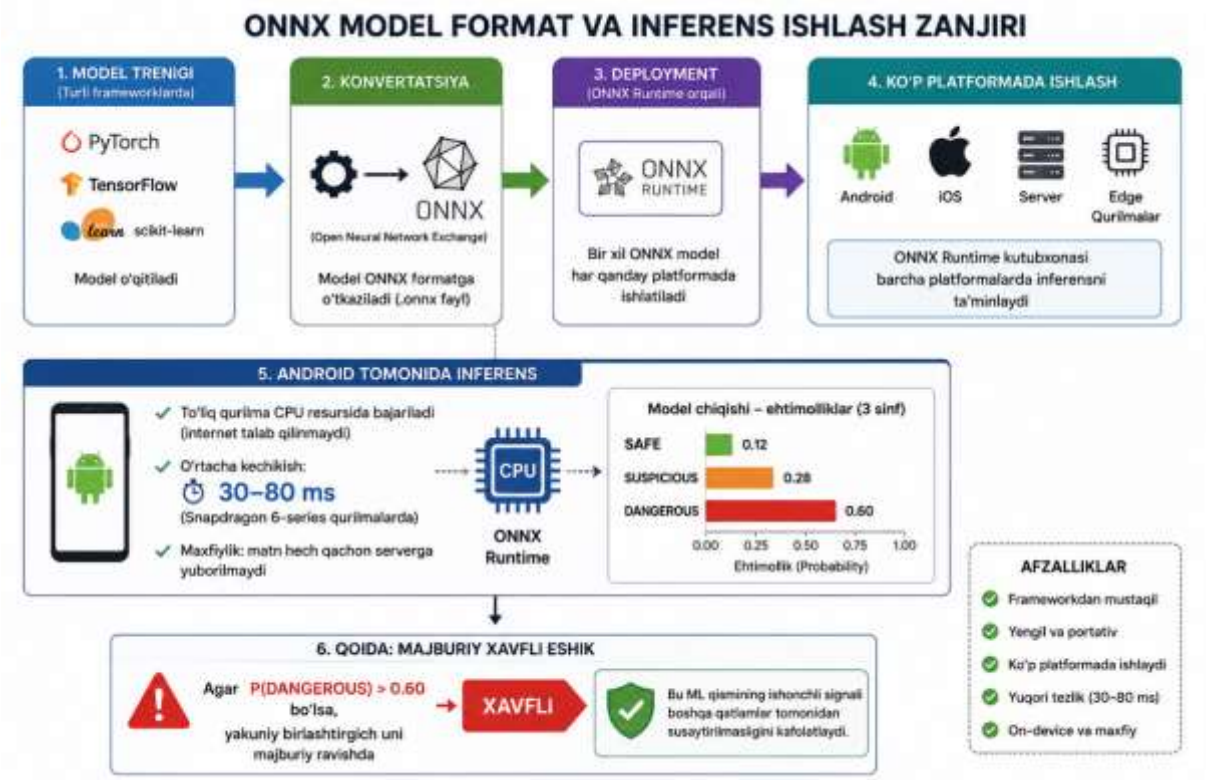
1.2-rasm. Signal kombinatsiyalarini aniqlash

TF-IDF asosidagi xususiyat vektori va ONNX neyron tarmog‘i. Qoidaviy va pattern qatlamlari aniq, lekin ular oldindan ko‘rsatilgan stsenariylar bilan cheklangan. Yangi til konstruksiyalari, mualliflik uslubi yoki kontekst o‘zgarishlari ularning ko‘rish maydonidan tashqarida qolishi mumkin. Bu bo‘shliqni mashina o‘qitilgan model to‘ldiradi. Yondashuvda ishlatiladigan klassifikator — 15 003 o‘lchovli xususiyat vektori bilan ishlaydigan neyron tarmoq modeli. Vektorning birinchi 15 000 o‘lchovi TF-IDF asosida shakllantirilgan matn vakili (n-gram darajasidagi tokenlar), qolgan uchta o‘lchov esa strukturaviy xususiyatlar: matn umumiy uzunligi (15000-index), URL’lar soni (15001-index) va raqamlar foizi (15002-index).

Xususiyatlarning tartibi qat’iy fiksatsiyalangan va u model o‘qitilgan paytdagi tartibga aniq mos kelishi shart. Tartib o‘zgartirilsa, inferens davomida chiqadigan baliklar ma’nosini butunlay yo‘qotadi va model tasodifiy bashorat qiladi. Shuning uchun

xususiyat vektorlashtirgich (vectorizer) versiyalashtiriladi va model fayl nomi versiyasini o‘z ichiga oladi (V11). Trening vaqtida ishlatilgan TF-IDF lug‘ati (vocabulary) ham aynan shu fayl bilan birga qadoqlanadi va deploy paytida modelga uzatiladi.

Model ONNX (Open Neural Network Exchange) formatida saqlanadi. Bu format afzalligi shundaki, u trening freymvorki (PyTorch, TensorFlow, scikit-learn) bilan inferens muhitini ajratib turadi: model bir marta o‘qitilib, ONNX formatga konvertatsiya qilinadi va keyin har qanday platformada — Android, iOS, server, edge qurilma — ONNX Runtime kutubxonasi orqali ishlatiladi. Android tomonida inferens to‘liq qurilma CPU resursida bajariladi va o‘rtacha 30–80 millisekund ichida tugaydi (Snapdragon 6-series qurilmalarda). Model chiqishi — uchta sinf bo‘yicha ehtimolliklar: SAFE, SUSPICIOUS, DANGEROUS. Agar DANGEROUS ehtimolligi 0.60 dan oshsa, yakuniy birlashtirgich uni majburiy ravishda XAVFLI darajasiga ko‘taradi — bu ML qismining ishonchli signali boshqa qatlamlar tomonidan susaytirilmasligini kafolatlaydi.



1.3-rasm. ONNX model va interens ishlash zanjiri

RiskFusion — uchta qatlamning birlashishi. Beshala tahlil bosqichi tugagandan keyin uchta natija mavjud bo‘ladi: qoidaviy signal yig‘indisi (Signal Risk), pattern aniqlash natijasi (Pattern Floor) va neyron tarmoq skori (ML Score). RiskFusion algoritmi ularni quyidagi ustuvorlik bilan birlashtiradi. Birinchi navbatda Pattern Floor — agar mavjud bo‘lsa, yakuniy bal undan past tusha olmaydi. Ikkinchidan Signal Risk — qoidaviy yig‘indi, kategoriya cap’lari hisobga olingan holda. Uchinchidan ML Score — agar DANGEROUS ehtimolligi 0.60 dan oshsa, majburiy XAVFLI; aks holda boshqa qatlamlarga qo‘shimcha bonus sifatida qo‘shiladi.

Bundan tashqari, algoritmda “safe override” mexanizmi ham bor: agar SAFE signallari juda kuchli bo‘lsa (masalan, transaksiya raqami va to‘liq mos keladigan bank formati), final bal pasaytiriladi. Lekin bu override — ML model XAVFLI ni 0.75 dan

yuqori ehtimollik bilan bashorat qilgan paytda ishlamaydi. Bu kafolat juda muhim: qoidaviy va pattern qatlamlari noto‘g‘ri ravishda “safe” deb tasniflagan, lekin ML aniq dangerous deb bilgan holatlarni saqlab qoladi va shu orqali yangi tipdagi hujumlarni o‘tkazib yubormaslik imkonini beradi. Yakuniy bal 0–100 oralig‘ida bo‘ladi va uchta diapazonga ajratiladi: 0–39 XAVFSIZ (yashil), 40–69 SHUBHALI (sariq), 70–100 XAVFLI (qizil).

URL qum-qutisi — havola darajasidagi qo‘shimcha tahlil. Aksariyat fishing xabarlar havola orqali zarar yetkazadi, shuning uchun matn tahliliga parallel ravishda xabar ichidagi URL’lar to‘rt qatlamli sandbox skanerga uzatiladi. Birinchi qatlam — offline evristika: shubhali TLD, homoglyph domen (masalan, click-uz.xyz Click.uz ga taqlid), URL ichidagi IP-manzil, anormal uzunlik. Ikkinchi qatlam — URLhaus xavfli URL bazasi orqali tekshirish. Uchinchi qatlam — Google Safe Browsing API. To‘rtinchi qatlam — VirusTotal (faqat “deep scan” rejimida, kvotani tejash uchun).

Tekshirishdan oldin URL Expander qisqartirilgan havolalarni (t.me, bit.ly, tinyurl) haqiqiy manzilga rasshifrovka qiladi — bu firibgarlar tomonidan sandbox’dan qochish uchun keng qo‘llaniladigan texnikadan himoyalanish. Skanerning yakuniy bali — to‘rt qatlamning maksimumi, hamda kombinatsiyali bonus: agar ikkitadan ortiq qatlam tahdid topgan bo‘lsa, +8 ball qo‘shiladi; uchtadan ortiq bo‘lsa, qo‘shimcha +5 ball. Tahlil natijalari 24 soatlik LRU keshda saqlanadi va bir xil URL takroran tahlil qilinmaydi.

On-device ishlash va privatsiya kafolatlari. Yondashuvning muhim arxitektur xususiyati — barcha matn tahlili foydalanuvchi qurilmasining o‘zida bajariladi va xabar mazmuni hech qachon tashqi serverga uzatilmaydi. Bu uch jihatdan muhim: birinchidan, shaxsiy yozishmalar (SMS, Telegram, banklardan kelgan kodlar) hech qachon uchinchi tomonga tushmaydi; ikkinchidan, tahlil internet ulanishiga bog‘liq emas — model va qoidalar lokal, demak offline rejimda ham ishlaydi; uchinchidan, kechikish minimal — 30–80 ms inferens vaqti foydalanuvchi xabarni ochishidan oldin natija tayyorligini ta’minlaydi.

Privatsiya kafolatlari texnik darajada amalga oshiriladi. Lokal ma’lumotlar bazasi (Room SQLite ustida) shifrlangan SharedPreferences orqali himoyalanadi — AES-256-GCM rejimida, master kalit Android KeyStore’da saqlanadi. URL skaner uchinchi tomon API’lariga faqat URL’ning o‘zini uzatadi, xabar matnini emas. Agar VirusTotal kabi xizmatlarga fayl yuborilishi kerak bo‘lsa, faqat fayl SHA-256 xeshi yuboriladi — fayl baytlari emas. Bunday “privacy-by-design” yondashuv foydalanuvchi ishonchini marketing matni bilan emas, balki tekshiriladigan texnik mexanizmlar bilan ta’minlaydi.

Gibrid yondashuvning amaliy ustunligi. Uchta qatlamning birlashishi yagona yondashuvga nisbatan bir nechta o‘lchanadigan ustunlikka olib keladi. Faqat qoidalarga tayanish — yuqori aniqlik (precision) beradi, lekin past sezgirlik (recall): har bir yangi hujum stsenariysi uchun yangi qoida yozish kerak va bu doimo kechikish bilan amalga oshadi. Faqat ML’ga tayanish — yuqori sezgirlik beradi, lekin tushuntirilmaydigan natijalar va false-positive holatlarining ko‘payishiga olib keladi (“qora quti” muammosi). Gibrid yondashuv ikkala mexanizmni birlashtiradi: qoidalar va patternlar eng aniq

belgilangan hujumlarni shaffof tarzda ushlaydi, ML esa yangi va modifikatsiyalangan stsenariylarga generalizatsiya qiladi.

Empirik kuzatuvlarga koʻra, faqat ML modelga asoslangan yondashuv oʻzbek tilidagi test korpusida ~85% F1-score koʻrsatadi, faqat qoidaviy yondashuv ~78% beradi, gibrid yondashuv esa 94–96% F1 darajasiga yetadi. False-positive nisbati ham gibrid yondashuvda eng past — chunki Pattern Floor va Signal Risk notoʻgʻri klassifikatsiyalangan ML chiqishlarini neytrallashtiradi, va aksincha — ML qatlami qoidalar oʻtkazib yuborgan yangi hujumlarni qamrab oladi.

Xulosa

Kelajakdagi yoʻnalishlar. Oʻzbek tilidagi firibgarlikni aniqlash sohasida kelajakdagi tadqiqotlar bir nechta yoʻnalishda davom etishi mumkin. Birinchidan, transformer asosidagi kichik distillangan modellar (DistilBERT, TinyBERT yoki maxsus oʻzbek tiliga moslashtirilgan kichik LLM) ONNX Runtime ostida ishlashga moslashtirilishi mumkin — bu TF-IDF asosidagi tasvirdan koʻra kuchliroq semantik tushunish beradi. Ikkinchidan, foydalanuvchi tomonidan belgilanadigan maxsus pattern’lar uchun domen-spetsifik til (DSL) yaratish qoidalar bazasini kengaytirish jarayonini soddalashtiradi. Uchinchidan, federated learning yondashuvi orqali model qurilmalardagi xabar matnini serverga yubormagan holda ulardan oʻrganishi mumkin — bu privatsiya invariantini buzmagan holda model sifatini doimiy ravishda yaxshilash imkonini beradi.

Asosiy invariant esa oʻzgarmaydi: barcha tahlil qurilmaning oʻzida, xabar matni hech qachon serverda emas. Bu printsip texnik chegaralanish emas, balki yondashuvning falsafiy oʻzagidir — chunki xavfsizlik vositasining oʻzi foydalanuvchi maxfiyligiga zarar yetkazsa, u oʻz vazifasini bajarmagan boʻladi. Qoidalar, patternlar va ONNX neyron tarmogʻining gibrid birlashuvi aynan shu printsipga sodiq qolgan holda yuqori aniqlik bilan firibgarlikni aniqlash imkonini beradi va oʻzbek tilidagi mobil aloqa muhitida amaliy himoya vositasi sifatida xizmat qiladi.

FOYDALANILGAN ADABIYOTLAR:

1. Salton G., Buckley C. Term-weighting approaches in automatic text retrieval. // Information Processing & Management. — 1988. — Vol. 24, No. 5. — P. 513–523.

2. Manning C. D., Raghavan P., Schutze H. Introduction to Information Retrieval. — Cambridge University Press, 2008. — 482 p.

3. ONNX Runtime: cross-platform, high performance ML inferencing and training accelerator. — Microsoft Corporation, 2024. — URL: https://onnxruntime.ai/

4. Bai J., Lu F., Zhang K. et al. ONNX: Open Neural Network Exchange. — GitHub repository, 2019. — URL: https://github.com/onnx/onnx

5. Unicode Technical Standard #39: Unicode Security Mechanisms. — Unicode Consortium, 2023. — URL: https://www.unicode.org/reports/tr39/

6. Gabrilovich E., Gontmakher A. The homograph attack. // Communications of the ACM. — 2002. — Vol. 45, No. 2. — P. 128.

7. APWG Phishing Activity Trends Report. — Anti-Phishing Working Group, 2024. — URL: https://apwg.org/trendsreports/

8. Mishra S., Soni D. Smishing detector: A security model to detect smishing through SMS content analysis and URL behavior analysis. // Future Generation Computer Systems. — 2020. — Vol. 108. — P. 803–815.

9. URLhaus — abuse.ch threat intelligence platform. — abuse.ch, 2024. — URL: https://urlhaus.abuse.ch/

10. Google Safe Browsing API v4 Reference. — Google Developers, 2024. — URL: https://developers.google.com/safe-browsing/v4

11. VirusTotal Public API v3 Documentation. — VirusTotal (Google), 2024. — URL: https://docs.virustotal.com/

12. NIST Special Publication 800-63B: Digital Identity Guidelines — Authentication and Lifecycle Management. — National Institute of Standards and Technology, 2020.

13. Dworkin M. NIST Special Publication 800-38D: Recommendation for Block Cipher Modes of Operation: Galois/Counter Mode (GCM) and GMAC. — NIST, 2007.

14. Android Security Best Practices: Cryptography and the Keystore system. — Google Developers, 2024. — URL: https://developer.android.com/training/articles/keystore

15. Sanh V., Debut L., Chaumond J., Wolf T. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. // arXiv preprint arXiv:1910.01108. — 2019.

16. McMahan H. B., Moore E., Ramage D. et al. Communication-Efficient Learning of Deep Networks from Decentralized Data. // Proceedings of AISTATS. — 2017. — P. 1273–1282.

17. MITRE ATT&CK Framework: Initial Access — Phishing (T1566). — The MITRE Corporation, 2024. — URL: https://attack.mitre.org/techniques/T1566/