GENERATIV AI YORDAMIDA TEXNIK HUJJATLARNI AVTOMATIK YARATISH TEXNOLOGIYALARI

Mamirov Diyorbek Rasiljon o‘gli[1]
[1] Farg‘ona davlat texnika universiteti,
Kompyuter muhandisligi va sun’iy intellect kafedrasi
E-mail: d.r.mamirov@fstu.uz

Annotatsiya

Ushbu maqolada generativ sun’iy intellekt - xususan, katta til modellari (LLM) va retrieval-augmented generation (RAG) - texnik hujjat yaratish jarayonida qo'llash texnologiyalari tadqiq etilgan. Mavjud yondashuvlar - shablonga asoslangan, qoidaga asoslangan va LLM-asoslangan - tizimli taqqoslama tahlil qilindi; GPT- 4o, Claude 3 Sonnet va Mistral-7B modellarining 360 ta real texnik hujjatlar to'plamida (API, foydalanuvchi qo‘llanmasi, texnik tavsif) ishlashi baholandi. RAG bilan boyitilgan GPT-4o tizimi an’anaviy shablon asosidagi yondashuvga nisbatan ROUGE-L ko‘rsatkichini 0,38 dan 0,71 ga (+86,8%), BERTScore-F1 ni 0,79 dan 0,91 ga (+15,2%), hujjat to‘liqligi darajasini esa 61,4% dan 87,6% ga oshirdi. Shu bilan birga gallyutsinatsiya darajasi 14,2% dan 3,8% ga pasaydi. Natijalar generativ AI ning texnik hujjatlashtirishni tubdan o‘zgartirish salohiyatini ko'rsatmoqda va O‘zbekiston dasturiy ta’minot sohasida tatbiq etish yo‘nalishlari taklif etilmoqda.

Kalit so‘zlar: *generativ AI, LLM, texnik hujjat, RAG, GPT-4o, ROUGE, BERTScore, avtomatik hujjatlashtirish, dasturiy ta'minot muhandisligi, prompt muhandislik.*

KIRISH

Dasturiy ta’minot muhandisligi amaliyotida texnik hujjat yaratish - API referencelari, foydalanuvchi qo‘llanmalari, arxitektura tavsiflari, o‘rnatish ko‘rsatmalari - ishlab chiqish jarayonining muhim, ammo ko‘pincha e’tibordan chetda qoladigan qismi hisoblanadi. Stack Overflow Developer Survey 2024 ma’lumotlariga ko‘ra [1], dasturchilarning 68,3% eskirgan yoki to‘liq bo‘lmagan hujjatni dasturiy ta’minot loyihalarining asosiy muammolaridan biri deb ko‘rsatgan. GitHub tadqiqoti esa ochiq manba loyihalarining 43% da hujjat sifati past yoki mavjud emasligi aniqlagan [2]. Boshqa tomondan, yaxshi hujjatlashtirish dasturchining ish vaqtining 20–35% ini oladi [3].

Generativ sun’iy intellektning - xususan, katta til modellari (Large Language Models, LLM) ning - paydo bo‘lishi ushbu muammoni tubdan qayta ko‘rib chiqish imkonini berdi. GPT-4 [4], Claude 3 [5], Gemini [6] va ochiq manba Mistral [7], LLaMA [8] modellari nafaqat tabiiy til matnini, balki dasturiy koddan hujjat, hujjatdan kod skeleti, foydalanuvchi talab tavsifidan texnik spetsifikatsiya generatsiya qila oladi. 2023–2025 yillarda bu sohada GitHub Copilot Workspace, Mintlify, Swimlane, Notion AI, Confluence AI kabi ixtisoslashgan mahsulotlar paydo bo‘ldi.

Shu bilan birga, generativ AI ning texnik hujjatlashtirishdagi imkoniyatlari hali to‘liq baholanmagan: qaysi model qaysi hujjat turi uchun samaraliroq? RAG (Retrieval-Augmented Generation) qo‘shilishi qanchalik sifatni oshiradi? Gallyutsinatsiya - ya’ni faktik xato - naqadar katta muammo? Bu savollar ushbu tadqiqotning markaziy

muammosini tashkil etadi. Maqolaning ilmiy hissasi: (i) texnik hujjat yaratish uchun LLM-asoslangan tizimlarning to‘liq baholash metodologiyasi ishlab chiqildi; (ii) uch turdagi hujjat va uch model kombinatsiyasida (3×3 dizayn) eksperimental baholash o‘tkazildi; (iii) RAG arxitekturasini hujjat sifatiga ta’siri miqdoriy aniqlandi; (iv) O‘zbekiston dasturiy ta’minot sektori uchun amaliy yo‘l xaritasi taklif etildi.

1.1. Texnik hujjat turlari va ularning xususiyatlari

Texnik hujjat bir jins emas - u har xil auditoriya, maqsad va tuzilish talablariga ega. Ushbu tadqiqotda uchta asosiy tur o‘rganildi: (1) API Documentation - dasturchilar uchun funksiyalar, parametrlar, qaytarish qiymatlari va kod namunalari; tuzilishi juda qat’iy (OpenAPI/Swagger standart); (2) User Manual - oxirgi foydalanuvchilar uchun vazifaga yo‘naltirilgan qo‘llanma; aniqlik va oddiylik muhim; (3) Technical Specification - muhandislar va menejerlar uchun tizim talablari, arxitektura va ishlash ko‘rsatkichlari; to‘liqlik va aniqlik birinchi o‘rinda. Har bir tur uchun LLM ning ustunliklari va kamchiliklari farqli - bu tezislashtirilgan tasnif muqaddimasi kelajakdagi natijalar talqinini osonlashtiradi.

2. MATERIALLAR VA METODLAR

2.1. Texnik hujjat generatsiyasida LLM qo‘llash arxitekturasi

Ushbu tadqiqotda generativ AI asosidagi texnik hujjat yaratish tizimining umumlashtirilgan besh qatlamli arxitekturasi taklif etiladi:

1-jadval.

Generativ AI asosidagi texnik hujjat tizimining besh qatlamli arxitekturasi

Qatlam	Komponent	Funksiya	Texnologiya/standarti
1. Ma’lumot kiritish	Kod, API spetsifikatsiya, talablar tavsifi	Birlamchi manba sifatida strukturalash	Python AST, OpenAPI 3.0, Markdown
2. Kontekst boyitish (RAG)	Vektor ma’lumotlar bazasi + semantik qidiruv	Domen-xos bilimlarni qo‘shish	Chroma/Pinecone + text-embedding-3-large
3. Prompt muhandislik	Tizim so‘rovi, bir nechta namuna, format ko‘rsatma	LLM chiqishini nazorat qilish	Chain-of-Thought, Few-shot, CoSTAR
4. Generatsiya	LLM yadro generatsiyasi	Hujjat matnini ishlab chiqish	GPT-4o / Claude 3 / Mistral-7B
5. Post-prosessng	Tekshiruv, formatlash, fayl chiqarish	Sifat nazorati va yakun shakl	LangChain, Pandoc, Markdownlint

Izoh: *RAG qatlami ixtiyoriy, ammo domen-xos loyihalarda (xususiy API, maxsus terminologiya) sezilarli sifat oshishini ta’minlaydi.*

Prompt muhandislik strategiyasi. Ushbu tadqiqotda CoSTAR (Context, Objective, Style, Tone, Audience, Response) tuzilishi [9] asosida standartlashtirilgan prompt shablon ishlatildi. API Documentation uchun tizim so‘rovida: «Sen tajribali texnik muharrirsan. OpenAPI 3.0 spetsifikatsiyasi bo‘yicha dasturchi auditoriyasi uchun to‘liq, aniq va kod namunali API hujjatini yarat. Har bir endpoint uchun: maqsad, parametrlar (tur, tavsif, majburiy/ixtiyoriy), javob kodlari (200, 400, 401, 404, 500), curl va Python kod misoli.» Bu tizim so‘rovi 12 ta iteratsiya orqali pilotlashtirildi va eng yuqori ROUGE-L ko‘rsatkichi beradigan variant tanlandi.

RAG tizimining tuzilishi. Retrieval-Augmented Generation komponenti uchta bosqichda ishlaydi: (i) Indekslash - mavjud korporativ hujjatlar, API standartlari va domain glossariy text-embedding-3-large modeli (1536 o‘lchamli vektor) orqali kodlanib, ChromaDB ga saqlanadi; (ii) Qidiruv - foydalanuvchi so‘roviga mos top-5 bo‘lim semantik o‘xshashlik asosida tanlanadi (cosine similarity ≥ 0,75); (iii) Kontekst qo‘shish - tanlangan bo‘limlar LLM prompt ning «kontekst» qismiga joylashtiriladi. RAG qo‘shilishi LLM «xotiradan uydirish» hodisasini kamaytirib, spetsifik domen bilimlarini ishonchli manba asosida generatsiya qilishga imkon beradi.

2.2. Baholash ma’lumotlar to‘plami

Eksperimental baholash uchun 360 ta real texnik hujjat namunasidan iborat to‘plam tuzildi. Ma’lumotlar uchta manbadan: GitHub ochiq manba loyihalari (Apache-2.0 litsenziya) - 180 ta namuna; O‘zbekiston IT kompaniyalari bilan hamkorlik (anonimlashtirish bitimi asosida) - 120 ta namuna; Stack Overflow Documentation arxivi - 60 ta namuna. Hujjat turlari bo‘yicha: API Documentation - 120 ta (33,3%); User Manual - 120 ta (33,3%); Technical Specification - 120 ta (33,3%). Har bir hujjat uchun: original manba (kod/talablar), etalon hujjat (reference document), model generatsiyasi saqlanib, baholash matritsasi to‘ldirildi.

2-jadval.

Baholash ma’lumotlar to‘plami tarkibi

Hujjat turi	Namunalar soni	O‘rtacha hajm (so‘z)	Murakkablik darajasi	Asosiy manba tili
API Documentation	120	1 840 ± 420	Yuqori (texnik aniqlik)	Ingliz + O‘zbek (20%)
User Manual	120	3 260 ± 680	O‘rta (soddalik muhim)	O‘zbek + Ingliz (40%)
Technical Specification	120	4 120 ± 920	Juda yuqori (to‘liqlik)	Ingliz (80%)
Jami	**360**	3 073 ± 780 (o‘rtacha)	-	-

Izoh: *O‘zbek tilidagi namunalar asosan foydalanuvchi qo‘llanmalari kategoriyasida ko‘proq - bu milliy IT bozori xususiyatini aks ettiradi.*

2.3. Baholash ko‘rsatkichlari

Hujjat sifatini baholash uchun uchta avtomatik va ikkita ekspert ko‘rsatkichi qo‘llanildi:

3-jadval.

Texnik hujjat sifatini baholash ko‘rsatkichlari tizimi

Ko‘rsatkich	Tur	Formula / Hisoblash	Interpretatsiya
ROUGE-L	Avtomatik	F1 asosida eng uzun umumiy ketma-ketlik	0-1; 0,70+ - yuqori; 0,50–0,70 - o‘rta
BERTScore-F1	Avtomatik	BERT embedding‘lar kosinusli o‘xshashligi	0-1; 0,90+ - juda yuqori; 0,85+ - yaxshi
BLEU-4	Avtomatik	4-gram aniqlik geometrik o‘rtachasi	0-1; texnik hujjat uchun 0,40+ qabul qilinadi
Hujjat to‘liqligi (DC)	Ekspert	DC = Mavjud_bo‘lim / Talab_bo‘lim × 100%	100% - to‘liq; 80%+ - qabul qilinadi
Gallyutsinatsiya darajasi (HR)	Ekspert	HR = Noto‘g‘ri_fakt / Barcha_fakt × 100%	< 5% - maqbul; < 2% - industrial talab

Izoh: *Ekspert baholash uchun har bir namuna 2 ta mustaqil texnik mutaxassis tomonidan ko‘rib chiqildi (Cohen’s κ = 0,82 - yuqori kelishuv).*

2.4. Sinovdan o‘tkazilgan tizimlar

Uchta LLM va ikkita arxitektura varianti kombinatsiyasi (3×2 = 6 konfiguratsiya) baholandi. Modellar: (A) GPT-4o (OpenAI, 128K kontekst, function calling); (B) Claude 3 Sonnet (Anthropic, 200K kontekst, constitution AI); (C) Mistral-7B-Instruct-v0.2 (ochiq manba, 32K kontekst, mahalliy joylashtirish). Arxitektura: (1) Vanilla - faqat LLM, RAG siz; (2) RAG - LLM + ChromaDB + text-embedding-3-large. Barcha sinovlarda temperature = 0,1, max_tokens = 4096. Har bir 360 ta namuna 2 marta ishlatildi (takrorlanish uchun) va o‘rtacha qiymat olinди.

3. NATIJALAR

3.1. Umumiy ishlash ko‘rsatkichlari

360 ta namunada barcha 6 ta konfiguratsiyaning ishlash natijalari:

4-jadval.

Barcha konfiguratsiyalar bo‘yicha qiyosiy baholash natijalari (n=360)

Konfiguratsiya	ROUGE-L	BERTScore-F1	BLEU-4	DC (%)	HR (%)	Tezlik (so‘z/s)
Shablon asosli (baseline)	0,38 ± 0,04	0,79 ± 0,03	0,22 ± 0,03	61,4 ± 4,2	n/a	-
GPT-4o Vanilla	0,61 ± 0,05	0,87 ± 0,03	0,41 ± 0,04	78,3 ± 3,8	14,2 ± 2,1	87
GPT-4o + RAG	**0,71 ± 0,04**	**0,91 ± 0,02**	**0,54 ± 0,04**	**87,6 ± 3,1**	**3,8 ± 0,9**	74
Claude 3 Sonnet Vanilla	0,59 ± 0,05	0,86 ± 0,03	0,39 ± 0,04	76,1 ± 4,1	15,7 ± 2,3	112
Claude 3 Sonnet + RAG	0,68 ± 0,04	0,90 ± 0,02	0,51 ± 0,04	84,9 ± 3,3	4,6 ± 1,1	98
Mistral-7B Vanilla	0,44 ± 0,06	0,82 ± 0,04	0,28 ± 0,05	65,2 ± 4,8	22,4 ± 3,1	184
Mistral-7B + RAG	0,56 ± 0,05	0,86 ± 0,03	0,38 ± 0,05	75,8 ± 4,0	9,2 ± 1,8	161

Izoh: *± 95% ishonch intervali; Wilcoxon signed-rank test: barcha GPT-4o+RAG vs baseline farqlari $p < 0,001$; DC - hujjat to‘liqligi; HR - gallyutsinatsiya darajasi; n/a - shablon tizimida HR o‘lchanmadi (statik chiqish).*

3.2. Hujjat turi bo‘yicha tahlil

Uchta hujjat turi bo‘yicha GPT-4o + RAG konfiguratsiyasining batafsil natijalari:

5-jadval.

GPT-4o + RAG konfiguratsiyasi natijalari hujjat turi bo‘yicha

Hujjat turi	ROUGE-L	BERTScore-F1	DC (%)	HR (%)	Asosiy kuchli tomoni	Asosiy zaif tomoni
API Documentation	0,74 ± 0,04	0,93 ± 0,02	91,2 ± 2,8	2,4 ± 0,7	Endpoint tavsifi, parametrlar	Murakkab nested ob’ektlar
User Manual	0,68 ± 0,05	0,89 ± 0,03	84,7 ± 3,4	4,1 ± 1,2	Vazifaga yo‘naltirilgan tuzilish	O‘zbek tilida grammatika
Technical Specification	0,71 ± 0,05	0,91 ± 0,02	86,9 ± 3,1	4,9 ± 1,4	Talablar seksiyalari	Ishlash ko‘rsatkichlari aniqlik

Izoh: *API Documentation da eng yuqori HR=2,4% - bu industrial talabni (< 5%) qondiradi va deyarli < 2% (juda yaxshi) ga yaqin.*

API Documentation da eng yaxshi natijalarning sababi: API hujjati OpenAPI/Swagger sxemasi kabi qat’iy tuzilishga ega - model inputdagi sxemani to‘g’ridan-to‘g’ri hujjat seksiyalariga moslashtira oladi. RAG tizimi mavjud endpoint tavsiflarini va kod namunalari to‘plamini kontekst sifatida ta’minlaydi. Aksincha, User Manual da o‘zbek tilidagi grammatik xatolar (fe’l tuslanishi, so‘z tartibi) va Technical Specification da ishlash ko‘rsatkichlari (ms, RPS, qo‘shimcha xarajat) ni to‘g’ri qadrlash muammolari qoldi.

3.3. RAG ta’sirini miqdoriy tahlil qilish

RAG qo‘shilishining har bir model uchun ta’siri:

6-jadval.

RAG arxitekturasining sifat ko‘rsatkichlariga qo‘shgan hissasi (Δ - Vanilla dan o‘zgarish)

Model	ΔROUGE-L	ΔBERTScore	ΔDC (pp)	ΔHR (pp)	RAG indeks samarasi (RI)
GPT-4o	+0,10 (+16,4%)	+0,04 (+4,6%)	+9,3 pp	−10,4 pp	**1,000 (etalon)**
Claude 3 Sonnet	+0,09 (+15,3%)	+0,04 (+4,7%)	+8,8 pp	−11,1 pp	0,942
Mistral-7B	+0,12 (+27,3%)	+0,04 (+4,9%)	+10,6 pp	−13,2 pp	0,871

Izoh*: RI (RAG Index) = RAG bilan sifat / GPT-4o+RAG sifati, barcha ko‘rsatkichlar bo‘yicha o‘rtacha. Mistral-7B ning nisbiy ΔROUGE (+27,3%) GPT-4o dan yuqori - bu kichik model uchun RAG ning muhimligini ta’kidlaydi.*

Muhim xulosa: RAG ta’siri modelning boshlang’ich kuchiga teskari proporsional. Mistral-7B (kuchsiz model) RAG dan nisbatan ko‘proq foyda ko‘radi (+27,3% ROUGE-L), GPT-4o esa (+16,4%) - ya’ni RAG kuchsiz modellarga kuchliroq «yordam» beradi. Bu natija mahalliy joylashtirish (on-premise) uchun Mistral-7B + RAG kombinatsiyasining GPT-4o Vanilla dan arzonroq va raqobatbardosh alternativa sifatida amaliy ahamiyatini oshiradi: ROUGE-L 0,56 vs 0,61, biroq narxi 10–20 barobar past.

3.4. Gallyutsinatsiya hodisasini tahlil qilish

HR 14,2% (GPT-4o Vanilla) dan 3,8% (GPT-4o + RAG) ga tushishining mexanizmlari: RAG tizimi LLM ni haqiqiy manba bilan «boqlangan» holatda saqlaydi. Gallyutsinatsiya hodisalarini tasniflab, uchta asosiy tur aniqlandi: (1) Parametr xatosi - funksiya argumentlari noto‘g’ri ko‘rsatilishi (44% hodisalar, asosan Vanilla); (2)

Versiya chalkashuvi - eski API versiya ma’lumotlari yangi sifatida ko‘rsatilishi (28% hodisalar, RAG orqali sezilarli kamaydi); (3) Uydirma misol - mavjud bo‘lmagan endpoint yoki konfiguratsiya parametri chiqarish (28% hodisalar, RAG orqali ham 40% kamaydi). GPT-4o + RAG ning qolgan 3,8% HR asosan murakkab nested ob’ektlar va versiya o‘tish (migration) bo‘limlarida kuzatildi.

$$HR_reduction = (HR_vanilla - HR_RAG) / HR_vanilla \times 100\% = (14,2 - 3,8) / 14,2 \times 100\% \approx 73,2\%$$

4. MUHOKAMA

4.1. Natijalarning xalqaro adabiyotlar bilan qiyosi

Olingan ROUGE-L = 0,71 (GPT-4o + RAG) ko‘rsatkichi xalqaro adabiyotlar bilan solishtirildi. Chen va boshqalar [10] OpenAI Codex asosidagi API hujjat generatorida ROUGE-L = 0,64 ko‘rsatgan - bizning natijamiz +10,9% ustun. Bu farq asosiy sababi: (a) RAG arxitekturasini qo‘llash; (b) hujjat turi bo‘yicha ixtisoslashtirilgan prompt shablon. Shi va boshqalar [11] GPT-3.5 + few-shot o‘rganishda ROUGE-L = 0,58 ko‘rsatgan; Gao va boshqalar [12] RAG ga bag’ishlangan seminal maqolada fanlararo vazifalar uchun BERTScore + RAG ko‘rsatkichi 0,89 darajasida bo‘lgan - bizning BERTScore = 0,91 bu natijadan biroz yuqori, bu ixtisoslashtirilgan (hujjatlashtirish) domen uchun RAG samaradorligini тасдиklaydi.

HR = 3,8% ko‘rsatkichi industrial qo‘llash mezonlari bilan taqqoslaganda muhokamaga loyiq. Ji va boshqalar [13] keng ko‘lamli LLM gallyutsinatsiya sharhida tibbiy va yuridik sohalarda 2% dan past HR talab qilinishini, dasturiy ta’minot sohasida esa 5% chegarasi qabul qilinishini ko‘rsatgan. Bizning GPT-4o + RAG natijasi (3,8%) dasturiy ta’minot uchun maqbul (< 5%), ammo hali tijorat miqyosi uchun ideal emas (< 2%). Bu «so‘nggi tekshiruv» (final review) inson tomonidan amalga oshirilishi zarurligini ko‘rsatadi.

4.2. Prompt muhandislik va sifat aloqasi

Prompt sifatining chiqish sifatiga ta’sirini o‘lchash uchun ablation tadqiqoti o‘tkazildi: (i) Minimal prompt (faqat vazifa) → ROUGE-L = 0,51; (ii) Bir nechta namuna bilan (few-shot, 3 ta) → ROUGE-L = 0,62; (iii) CoSTAR tuzilishi bilan → ROUGE-L = 0,67; (iv) CoSTAR + RAG + format ko‘rsatma → ROUGE-L = 0,71. Har bir qo‘shimcha prompt komponenti sezilarli o‘sish berdi. Ayniqsa few-shot qo‘shish eng katta o‘sishni berdi (+21,6%) - bu LLM ning namunalar orqali format va uslubni tezda o‘zlashtirishini ko‘rsatadi. Bu natijadan amaliy xulosa: texnik hujjat generatsiyasida prompt muhandislikka investitsiya qilish RAG’dan ham kuchliroq «birinchi qadam» bo‘lishi mumkin.

4.3. Iqtisodiy tahlil: xarajat va foyda

Generativ AI asosidagi hujjat tizimining iqtisodiy asoslanishi uchun xarajat-foyda tahlili amalga oshirildi:

7-jadval.

Hujjat yaratish usullarining xarajat-foyda taqqoslama tahlili (1 ta o‘rtacha hujjatga)

Usul	Inson vaqti (soat)	Narxi (USD)	Sifat darajasi	Yangilash tezligi	Miqyoslanish
An’anaviy qo‘lda	6-12	60-180	Yuqori (ekspert)	Sekin (manual)	Cheklangan
Shablon asosli	2-4	20-60	O‘rta (standart)	O‘rta	O‘rta
GPT-4o Vanilla	0,3-0,8	2,4-8,6	Yaxshi	Tez (daqiqalar)	Juda yuqori
GPT-4o + RAG	0,4-1,0	3,2-11,4	**Juda yaxshi**	Tez (daqiqalar)	**Juda yuqori**
Mistral-7B + RAG (mahalliy)	0,5-1,2	0,8-2,4*	Yaxshi	Tez	Juda yuqori

Izoh: **Mistral-7B mahalliy joylashtirish xarajati GPU server ijarasini o‘z ichiga oladi (AWS g5.xlarge: $1,01/soat). 100+ hujjat/oy hajmida mahalliy joylashrish iqtisodiy ustunlik beradi.*

4.4. O‘zbekiston IT sektori uchun tatbiq yo‘l xaritasi

O‘zbekistonda IT sohasini rivojlantirish bo‘yicha Prezident Farmoni (2020) va IT Park O‘zbekiston faoliyatini hisobga olib [14], generativ AI asosidagi hujjatlashtirish tizimini joriy etishning ikki bosqichli yo‘l xaritasi taklif etiladi. Birinchi bosqich (2025–2026): IT Park rezident kompaniyalarida pilot loyiha sifatida GPT-4o + RAG tizimini ingliz tilidagi API hujjatlashtirishga joriy etish; ochiq manba to‘plamni O‘zbek tiliga kengaytirish. Ikkinchi bosqich (2026–2028): Mistral-7B yoki LLaMA-3 ni o‘zbek tilida fine-tuning qilish (O‘zbek NLP corpusi - 4 milliard token talab qilinadi [15]); davlat xizmatlari va moliya sektori hujjatlashtirish standartlari bilan integratsiya. Asosiy cheklov: O‘zbek tilidagi texnik hujjat korpusi juda kichik - bu muammoni hal qilish uchun UzText, MilliyCorpus kabi milliy loyihalarni rivojlantirish strategik zarurat.

5. XULOSA

Ushbu tadqiqotda generativ AI - xususan, GPT-4o va Claude 3 Sonnet kabi katta til modellari RAG arxitekturasi bilan birgalikda - texnik hujjat yaratishda an’anaviy shablon asosidagi va faqat LLM yondashuvlariga nisbatan barcha baholash mezonlari bo‘yicha statistik ahamiyatli va amaliy jihatdan muhim ustunlik ko‘rsatganligini 360 ta namunali eksperimental tadqiqot isbotladi.

Asosiy miqdoriy natijalar: ROUGE-L 0,38 → 0,71 (+86,8%); BERTScore-F1 0,79 → 0,91 (+15,2%); hujjat to‘liqligi (DC) 61,4% → 87,6% (+26,2 pp); gallyutsinatsiya darajasi (HR) 14,2% → 3,8% (−73,2%). API Documentation da HR = 2,4% industrial talabga (< 5%) mos keldi. RAG komponenti modellar bo‘yicha ROUGE-L ni o‘rtacha +19,7% oshirdi va HR ni −12,2 pp kamaytirdi. Mistral-7B + RAG

mahalliy joylashtirish uchun xarajat-sifat munosabati bo‘yicha eng samarali variant (GPT-4o + RAG sifatining 79% ini 7–10 barobor kam xarajatda ta’minlaydi) bo‘lib chiqdi.

Tadqiqot yana bir muhim xulosani tasdiqladi: generativ AI texnik hujjatni to‘liq almashtirishdan ko‘ra inson muharrirning ish unumdorligini keskin oshirishga qodir - manual hujjatlashtirish uchun 6–12 soat sarflanadigan bo‘lsa, AI yordamida 0,4–1,0 soatga (yakuniy tekshiruv bilan) tushади. O‘zbekiston IT sektori uchun taklif etilgan ikki bosqichli yo‘l xaritasi va mahalliy o‘zbek tili modelini rivojlantirish bo‘yicha tavsiyalar kelajak tadqiqotlar va sanoat-akademiya hamkorligining asosini tashkil etishi mumkin.

FOYDALANILGAN ADABIYOTLAR

[1] Stack Overflow. Developer Survey 2024. – 2024. URL: https://survey.stackoverflow.co/2024/.

[2] Zolduoarrati E., Karanikiotis T. Why developers struggle with documentation: An empirical study // IEEE International Conference on Software Maintenance and Evolution (ICSME). – 2021. – P. 412–422. DOI: 10.1109/ICSME52107.2021.00043.

[3] Forward A., Lethbridge T.C. The relevance of software documentation, tools and technologies: A survey // ACM Symposium on Document Engineering. – 2002. – P. 26–33.

[4] Achiam J., Adler S., Agarwal S. et al. GPT-4 Technical Report // arXiv preprint arXiv:2303.08774. – 2023. URL: https://arxiv.org/abs/2303.08774.

[5] Anthropic. Claude 3 Model Card. – 2024. URL: https://www.anthropic.com/claude-3-model-card.

[6] Team G., Anil R., Borgeaud S. et al. Gemini: A family of highly capable multimodal models // arXiv preprint arXiv:2312.11805. – 2023. URL: https://arxiv.org/abs/2312.11805.

[7] Jiang A.Q., Sablayrolles A., Mensch A. et al. Mistral 7B // arXiv preprint arXiv:2310.06825. – 2023. URL: https://arxiv.org/abs/2310.06825.

[8] Touvron H., Martin L., Stone K. et al. Llama 2: Open foundation and fine-tuned chat models // arXiv preprint arXiv:2307.09288. – 2023. URL: https://arxiv.org/abs/2307.09288.

[9] Schulhoff S., Ilie M., Balepur N. et al. The prompt report: A systematic survey of prompting techniques // arXiv preprint arXiv:2406.06608. – 2024.

[10] Chen M., Tworek J., Jun H. et al. Evaluating large language models trained on code // arXiv preprint arXiv:2107.03374. – 2021.

[11] Shi E., Schafer B., Baxter K. Automatically generating API documentation with large language models // ACM International Conference on the Foundations of Software Engineering (FSE). – 2024.

[12] Gao Y., Xiong Y., Gao X. et al. Retrieval-augmented generation for large language models: A survey // arXiv preprint arXiv:2312.10997. – 2023.

[13] Ji Z., Lee N., Frieske R. et al. Survey of hallucination in natural language generation // ACM Computing Surveys. – 2023. – Vol. 55, No. 12. – Art. 248. DOI: 10.1145/3571730.

[14] O‘zbekiston Respublikasi Prezidentining «Axborot texnologiyalari sohasini jadal rivojlantirish» to‘g’risidagi Farmoni. – Toshkent, 2020.

[15] Mansurov B., Mirzayev A. Building a large-scale Uzbek text corpus for NLP research // Proceedings of LREC-COLING. – 2024. – P. 1842–1849.

[16] Lin C.-Y. ROUGE: A package for automatic evaluation of summaries // ACL Workshop on Text Summarization Branches Out. – 2004. – P. 74–81.

[17] Zhang T., Kishore V., Wu F. et al. BERTScore: Evaluating text generation with BERT // International Conference on Learning Representations (ICLR). – 2020.

[18] Papineni K., Roukos S., Ward T., Zhu W.-J. BLEU: A method for automatic evaluation of machine translation // ACL. – 2002. – P. 311–318.