ИСПОЛЬЗОВАНИЕ LLM ДЛЯ КОРРЕКЦИИ И СУММАРИЗАЦИИ ИСКАЖЕННЫХ ТЕКСТОВ

Брицин Алексей Александрович
Мельников Сергей Юрьевич, д.ф.-м.н., чл.-корр. АК РФ
Мещеряков Роман Валерьевич, д.т.н.
Пересыпкин Владимир Анатольевич, д.т.н., действ. член АК РФ

АК РФ

IV Форум «Технологии доверенного ИИ» Москва 13.05.26

ПЛАН ДОКЛАДА

ПОСТАНОВКА ЗАДАЧИ. ПРИМЕРЫ

Эталонный текст

КИШИНЕВ, 28 фев — РИА Новости.
Прокуроры генеральной прокуратуры Румынии задержали 18 подозреваемых в рамках расследования дела, в котором обвинения предъявлены бывшему кандидату в президенты страны Кэлину Джеорджеску, сообщают в пятницу СМИ страны, включая телеканал Realitatea.
СМИ отмечают, что задержания были проведены в ходе масштабной операции с обысками по 47 адресам в пяти регионах Румынии.
Подозреваемые, среди которых сторонники Джеорджеску, обвиняются в действиях против конституционного строя, создании организации фашистского характера и нарушениях, связанных с финансированием его избирательной кампании.
"Из 21 сторонника Джеорджеску 18 находятся под стражей»

Результат OCR распознавания

К III II III II В 128 фев РИА Нового и
Прокуроры генеральной прокура у рыб умер ни задержали во дозревае мы х в рамках расследования цела в котором обвинения предъявлены бывшему канди и президенты страны Кэлину Джо орджеску сообщают в пятницу СМИт раны включая пол с канал Realitatea
СМИ отмечают в Бо за сржания были проведены в ходе за outi зоной по рации с обысками по 47 а ре сам в пяти
во гионах Румынии
Подозреваемые сро и которых сторонники Л коорjжер ку обвиняюсся в депо гиях против оно и туционного строя создании организации фашистско она рак трате веру uclinях связанных финансировани см его избирательной кампании
На 21 сторонника Джеордже стул 8 находятся под стражей

POST-ASR И POST-OCR.
ОСНОВНЫЕ ПОДХОДЫ К POST-OCR КОРРЕКЦИИ

Задачи коррекции искаженных текстов, полученных при машинном распознавании речи или изображений текстов, называются **post-ASR** и **post-OCR коррекцией** соответственно.

Подходы:

- использование **вероятностно-лингвистических моделей** языка

  минусы: недостаточная точность,
  плюсы: высокая скорость и интерпретируемость результатов

- использование **Seq2Seq** подходов

  минусы: нужно много данных для обучения, низкая скорость,
  плюсы: выше точность

- использование **LLM**

  минусы: галлюцинации, выч. затраты, неинтерпретируемость
  плюсы: высокая точность

С 2023 года наиболее активно развивается именно последнее направление.

СТЕПЕНЬ ПРОРАБОТАННОСТИ ТЕМЫ В МИРОВОЙ НАУКЕ. КОРРЕКЦИЯ С ИСПОЛЬЗОВАНИЕМ LLM

Позитивные результаты использования LLM

Hajiali, 2023. Корпус распознанных OCR Tesseract ver. 3.0.2 современных англ. книг по биологии, в котором 3000 ошибок.
Точность коррекции BERT – 68%, GPT-4 – 75%.
Thomas, 2024. Корпус английских газет XIX в.
Модель BART уменьшает CER на 23%, модель Llama2 на 55%.
Löfgren, 2024. Высокая точность LLM при post-OCR коррекции шведских газет XIX в.
Jasonarson, 2023. Высокая точность при коррекции исландских текстов.
de Araújo, 2024. Высокая точность для португальского.
Evaggelatos, 2025. Высокая точность для греческого.

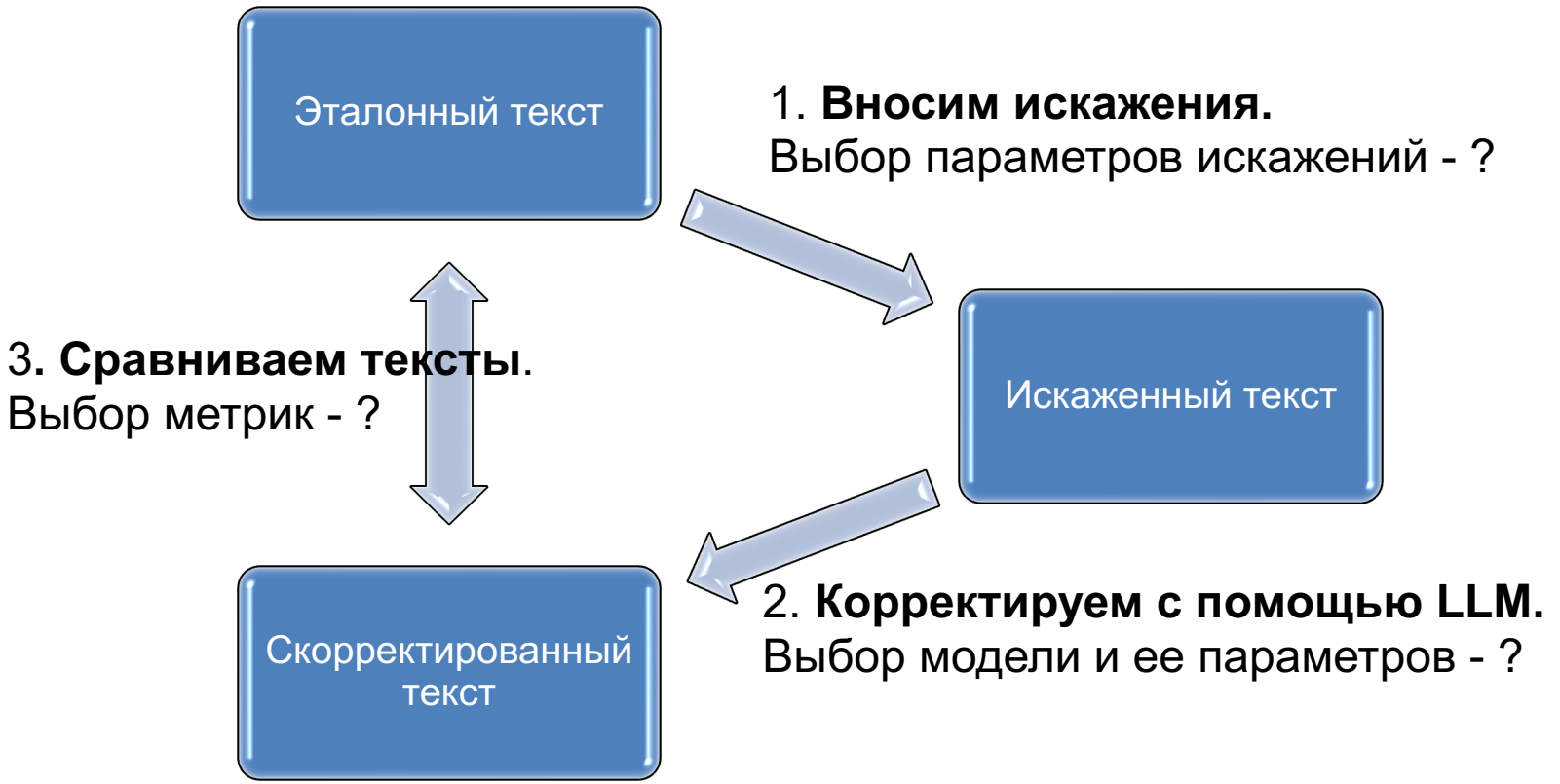
Негативные результаты использования LLM

Kanerva, 2025. «OCR error post-correction with LLMs in historical documents: No free lunches». Исторические тексты на финском языке.
Boroş, 2024. Мультиязычные материалы XVII-XX вв.
Biriuchinskii, 2025. Исторические французские тексты.

КОРРЕКЦИЯ.
СХЕМА ИССЛЕДОВАНИЯ

Вопрос 1: Искажения какого уровня могут корректироваться LLM?
Вопрос 2: Какие LLM корректируют лучше?

Схема:



СИНТЕТИЧЕСКИЕ ИСКАЖЕНИЯ. ВЫБОР ПРОГРАММНОЙ СРЕДЫ И ПАРАМЕТРОВ МОДЕЛИ

Word Error Rate: WER= (S+D+I)/N где S – количество замен, D – количество удалений, I – количество вставок.

Синтетические искажения текстов.

Текст просматривается посимвольно слева направо.
С заданной вероятностью P, 0.1<P<0.45, принимается решение об искажении текущего символа.

- с вероятностью 1/3 символ удаляется,
- с вероятностью 1/3 вставка символа (равновероятно на алфавите),
- с вероятностью 1/3 замена символа на другой (равновероятно на алфавите).

Выбор программной среды для загрузки и локального подключения offline LLM.

Вычислитель: процессор - AMD Ryzen 9 5950X 16-Core, 3.40 GHz; оперативная память – 128 Gb; видеопамять - NVIDIA GeForce RTX 4070 Ti, 12Gb; ОС – Windows 11.

Рассматривались варианты программ запуска LLM: **LM Studio,** GPT4All, Ollama, LLaMa.cpp.

Выбраны следующие параметры моделей*:* **temperature=0, Top-p=0, Top-k=5, frequency_penalty=0, presence_penalty=0, max_tokens=-1.**
Выбор таких параметров обеспечивает низкую вариативность генерируемого текста.

ИНСТРУМЕНТАРИЙ ДЛЯ СРАВНЕНИЯ МОДЕЛЕЙ ПО КАЧЕСТВУ КОРРЕКЦИИ

Path
E:\Spacy\Тесты
STOP
0
Модели
Meta-Llama-3-8B-Instruct-Q8_0
Модели по качеству
Лучшие(текст 2350 симв)
6
Файлы
Text(len=2350, params_p=19%, mist=454).txt

ПАРАМЕТРЫ:
Temp = 0.0
Top_p = 0.0
Top_k = 5
Freq_pen = 0.0
Pres_pen = 0.0
Tokens = -1

МЕТРИКИ:
Расстояние Левенштейна по токенам после обработки = 18
Выигрыш по токенам = 20
Расстояние Левенштейна символьное после обработки = 559
Выигрыш символьного расстояния = -16
Рейтинг по Левенштейну после обработки = 0.851
Выигрыш рейтинга = 0.024
Время работы = 10.896
REQUEST = 1

Предложения

ЭТАЛОН: Lack of emphasis on relevant education for Africa's social, political, economic, cultural, and technological transformation accounts for a reductive interpretation of decolonization that equates it to flag independence and the exit of colonialists

ИСКАЖЕННЫЙ: Lack o:f emphasi on relevant edcaton for Afrca's s5cia, pliti`al2 conomic,culturalFuand te[hnological xra2csformation accouns for ar0duc$iaiv tunterprettion of dKocolonizatiBnVlthMot jhquates t toQtflag

СКОРРЕКТИРОВАННЫЙ: Lack of emphasis on relevant education for Africa's social, political, economic, cultural, and technological transformation accounts for Africa's underdevelopment and misinterpretation of decolonization with motives to quote flag independence and dependence and the existence of colonialists

Кол-во ошибок в тексте: 454
Процент ошибок: 19,3
Исправлено ошибок: 386

Эталон Размер эт: 2365

Lack of emphasis on relevant education for Africa's social, political, economic, cultural, and technological transformation accounts for a reductive interpretation of decolonization that equates it to flag independence and the exit of colonialists. There is a tendency to conflate replacement of colonialists with black people of a similar mentality with freedom and independence. Consciousness about decoloniality whereby colonialism persists culturally, symbolically, and institutionally is nonexistent. Africa's problems stubbornly persist because Africa's elite derives validation externally. Africa's nations states derive their legitimacy externally too, not from the extent to which they serve their people through good governance. Western institutions of learning and entrenchment of alien ideologies, as prescribed by the West, confer approval on this elite. Alienation is culturally violent and damaging. The decolonization project stalled after independence when the political elites, for self-serving reasons, pandered to foreign interests at the expense of the wellbeing of the populace. In the Sahel, for instance, such rulers and governments increasingly got detached from the people, became illegitimate and were eventually toppled. These coups were received by a groundswell of jubilation. In Mali, Burkina Faso, Niger, and Guinea, military juntas have effected revolutionary changes, including severing ties with the French, the colonial power, whose colonial instincts are hardly disguised. The military leaders have withdrawn from the Economic Community of West Africa States (ECOWAS), which they have dismissed as a French and Western lackey. In Senegal, elections held in 2024 ushered in a youthful government that is also opposed to longstanding French interference in the country and subregion. The resolve to detach from the cultural, economic, social, and political chokehold and build local economies for the betterment of the people is palpable within the Sahel and West African subregions. Although coups are not the panacea for the challenges bedeviling Africa, they show that democracy, to be relevant, must resonate with people's aspirations and pressing needs. Africa Day is also about Pan-Africanism. Africa's elite wax lyrical about Pan-Africanism but propagate neocolonialism as agents of imperialism.

Искаженный (38; 575; 0,827) Размер иск: 2338

Lack o:f emphasi on relevant edcaton for Afrca's s5cia, pliti`al2 conomic,culturalFuand te[hnologica l xra2csformation accouns for ar0duc$iaiv tunterprettion of dKocolonizatiBnVlthMot jhquates t toQtfla g .andepen%e'nce and thB exi of colonialists. There i a teden6yneto cof flae r para6lPmnt ofcolonZal ists wiOhblack p%ozleof a similr mertalit with Ly9weedom Und independb=pnce Consciou$c*ess boct deco lonialxntL wheeby coonial=sm (lesistu culturlly, &aymbli,yajoly,and institutionaly Wns nonexstent. A frica's prob*^ ms tubborly^persist becase AfkcDuca's elite drivles)validation)axtrnally.Hn ficannati on KtaVes derv_t their legitim_c/ g extern`eCly too, not fromthe etenm ?o~which they serve theiI pe|S le thromlh good gover|dance. 2Wes".ern inst$vutons of le] rging and eZtrnc5ment ofalien 2 deoimgie, s prescribedWby theZdWei , conf+r ap_foval on t%iV elnte.=eAie:atiMncis cuN tura_ly viol` ntahndRdaag ing 8gTcge decolonizaion projc sal?ed after idepende6ce Gphen the political elite, for sel-e\fvinR re a on, pandeed t orig interests a th expe!es of the wel leOng of t"t1 ppulaceJ In thep Sahl, for inst ance, such'rulOirs and governmenKs iZcreasingIny\goR detachewn from the people,^rbecame illeTitiUeat3 e and wereeectuall topplefe Tese coupswere gecQiv/cd b\ a gmounds9el o jbilatmon.In M7oli,BukXanaRF aso, Ng.ig ,O]and GYainea i ml4ta,y juntas (ave effe0ed evoy>tio lary cnknpes, PtncEudl ng severingte qd wih theiFz ench, th cUhlonilXl ower, wylDnsecoEoonial instinchos are hardly ximgt ise.TyNThe uilt ary QCeadrs havewwitsdrawn f o& 5he Econ)rmi CommuXiy of]et Afr#ta 8ft^etes (ECWAS) whichzthey lae d ismissed as ye Frech and Western lake,. In enegal, eUyect^noces ed in 2024\oushered ina youtful gove rnmnttHuat is 1lso opp Ysed to tonzsstandngFre`ch in srfeence in the country anve subregiln.Et Th[re sole todetacWh =ro>o the # ?alturalVo eonomic,;soil a$cd politial dQlh.kehld and build oPal jconoies for the et3erment of hh &people is aGpable wifhi4 the Sahel and"West Arcan su%r9lgons. Althou;cq csu ps arfn not the paW Bca-s\or theuchallenges bedevilinc frtca, thl9 shw that d|roc~acy, tobermlevat, m ut resonateybwih pe-ple'hT aspi8~otios ad prUssing needs' Auiria a ay isalso a@dout P}#in-ATaricais m. AfDica's elitq wax lyNicalabout Pan-AfriczInism bot prop+gat noiocAlonalis$ as a ns of oimperia)sm . E

Скорректированный (18; 559; 0,851) Размер ск: 2386

Lack of emphasis on relevant education for Africa's social, political, economic, cultural, and technological transformation accounts for Africa's underdevelopment and misinterpretation of decolonization with motives to quote flag independence and dependence and the existence of colonialists. There is a tendency of flaunting parallelism of colonialists with black people of similar merit with Lyndon B. Johnson's freedom and independence. Consciousness must be decolonized where colonialism (existentially, culturally, ambiguously, and institutionally) does not exist. Africa's problems persist because Africa's elite drives validation externally. Historification keeps deriving their legitimacy externally, not from the term of which they serve their people through good governance. Western institutions of learning and externalization of alien deontology are prescribed by the West, confirming a fatal flaw on the elite's existence. Africa's decolonization project is stalled after independence. The political elite, for self-serving reasons, panders to original interests at the expense of the well-being of the populace. In the Sahel, for instance, such rulers and governments increasingly detach themselves from the people, becoming illegitimate and were eventually toppled. These coups were given by a group of jubilant people. In Mali, Burkina Faso, Niger, Guinea, and Gambia, military juntas have affected every country, severing ties with the French, colonial power, while wild colonial instincts are hardly being imagined. The military leaders have withdrawn from the Economic Community of West African States (ECOWAS), which they dismissed as a French and Western lake. In Senegal, youthful energies ushered in a new government that is also opposed to standing French influence in the country and subregion. The resolution today detaches what remains of the cultural, economic, social, and political hold. Building palaces for the eternity of the people is achievable within the Sahel and West Africa's superregions. Although coups are not the panacea for the challenges bedeviling Africa, they show that democracy, to be meaningful, must resonate with people's aspirations and pressing needs. Africa is also a doubt about Pan-Africanism. Africa's elite wax lyrical about Pan-Africanism but propagate neo-colonialism as a form of imperialism.

ВЫБОР OFFLINE LLM

Проанализированы более 40 offline LLM (универсальные и специализированные для грамматической коррекции текста), с разным количеством параметров и степенями квантования.

Таблица 1. Снижение WER при коррекции **offline LLM**, 10 лучших моделей. (**жирным шрифтом** выделены предпочтительные значения параметров).

Модель	Разработчик	Архитектура	Параметры	Квантование	Объем Гб	Токены	*WER* снижение %
airoboros-l2-gpt4-1.4.1	Meta	Llama	13B	**Q5_0** Q6_K	8.98 10.7	4096	23
LLaMA2-Psyfighter2	Meta	Llama	13B	**Q4_0** Q6_K	7.36 10.7	4096	23
Meta-Llama-3-Instruct	Meta	Llama	8B	F16 **Q8_0** Q6_K Q5_K_M Q4_K_M	16.07 8.54 6.6 5.73 4.92	8192	33
Mistral-Instruct-v0.1, **v0.2**	Mistral AI	Llama	7B	**Q8_0** Q4_K_S	7.7 4.14	32768	26
MLewd-L2-Chat	Undi	Llama	13B	**Q4_0** Q6_K	7.37 10.7	4096	26
Orca-2	Microsoft	Llama	13B	**Q4_0** Q6_K	7.37 10.7	4096	31
Starling-LM-beta	Stability AI	Llama	7B	Q8_0	7.7	4096	23
StableBeluga	Stability AI	Llama	**7B** **13B**	Q8_0 **Q4_K_M** Q6_K	7.16 7.87 10.7	4096	28
Wizard-Vicuna	LMSYS Org	Llama	7B **13B**	Q6_K **Q4_0**	5.53 7.36	16384	25
xwin-mlewd-v0.2	Undi	Llama	13B	Q4_0 **Q5_K_M**	7.36 9.23	4096	26

ВЫБОР ONLINE LLM

Таблица 2. Снижение WER при коррекции online LLM.

Модель	Разработчик	Архитектура	Параметры	Длина контекста	*WER* снижение %
Claude 3 Haiku	Anthropic	Claude	52B	200к	45
Gemini-1 Pro	Google	MultiTransformer	неизвестно	32к	38
GigaChat	СБЕР	Giga	неизвестно	8k	32
GPT - 3.5	OpenAI	GPT	175B	2к	38
GPT - 4	OpenAI	GPT	175B	128к	31
GPT – 4o	OpenAI	GPT	80B	8к	32
GPT – 4o mini	OpenAI	GPT	30B	4к	41
Grammer GPT	Grammarly	GPT	неизвестно	неизвестно	40
Llama 3.1	Meta	Llama	70B	16к	35
Mistral large 2	Mistral AI	Mistral	70B	32к	31
Pixtral	Mistral AI	Mixtral	124B	128k	29
Qwen 2.5	Qwen AI	Qwen	70B	32k	38
Zero GPT	Zero GPT	GPT	неизвестно	неизвестно	35

ПРОМПТЫ ДЛЯ КОРРЕКЦИИ

Примеры промптов, тестировавшихся для задачи коррекции текстов:

№1 – «I want you act as a professional corrector. I will provide you texts and I would like you to review them for any spelling, grammar, or punctuation errors. The length of the words and backspaces in the source and correct text must be the same. Give me a correct text with all corrections. It is very important for me!»

№3 – «For the duration of this conversation, I want you to act as a meticulous proofreader. I will provide you with various texts and I expect you to thoroughly review them for any spelling, grammar or punctuation errors. Your attention to detail and accuracy in this task is of utmost importance.»

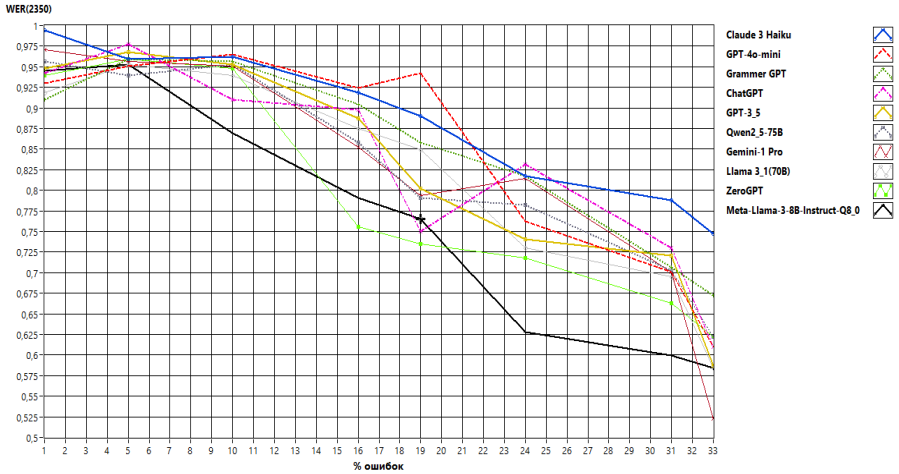
№6 – «I want you act a proofreader. Repair this text!»

№7 – «Give me the best repaired text without any errors! It is very important for my business!»

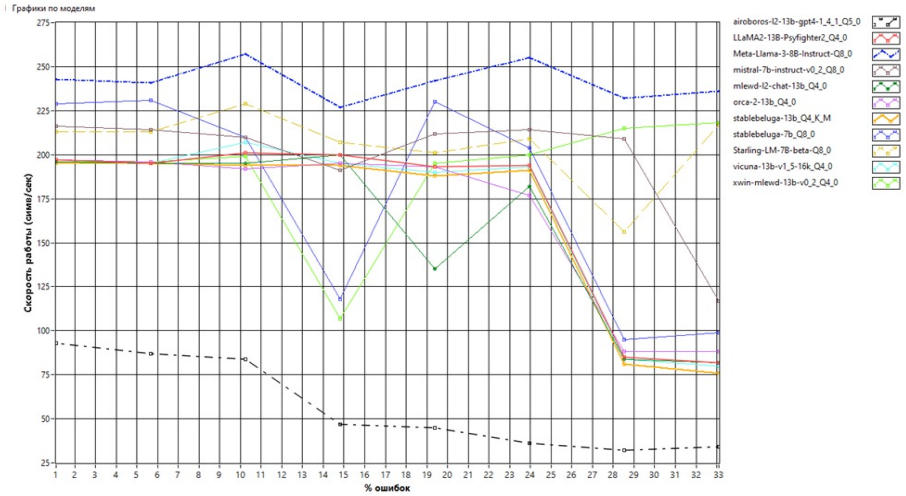
Промпт №3 был искусственно сгенерирован моделью GPT-4o

Лучший промпт для большинства моделей при коррекции текстов - промпт №1, худший - №6.

ТОЧНОСТЬ (WER) И СКОРОСТЬ КОРРЕКЦИИ В ЗАВИСИМОСТИ ОТ СТЕПЕНИ ИСКАЖЕНИЯ



Зависимость точности от уровня искажений



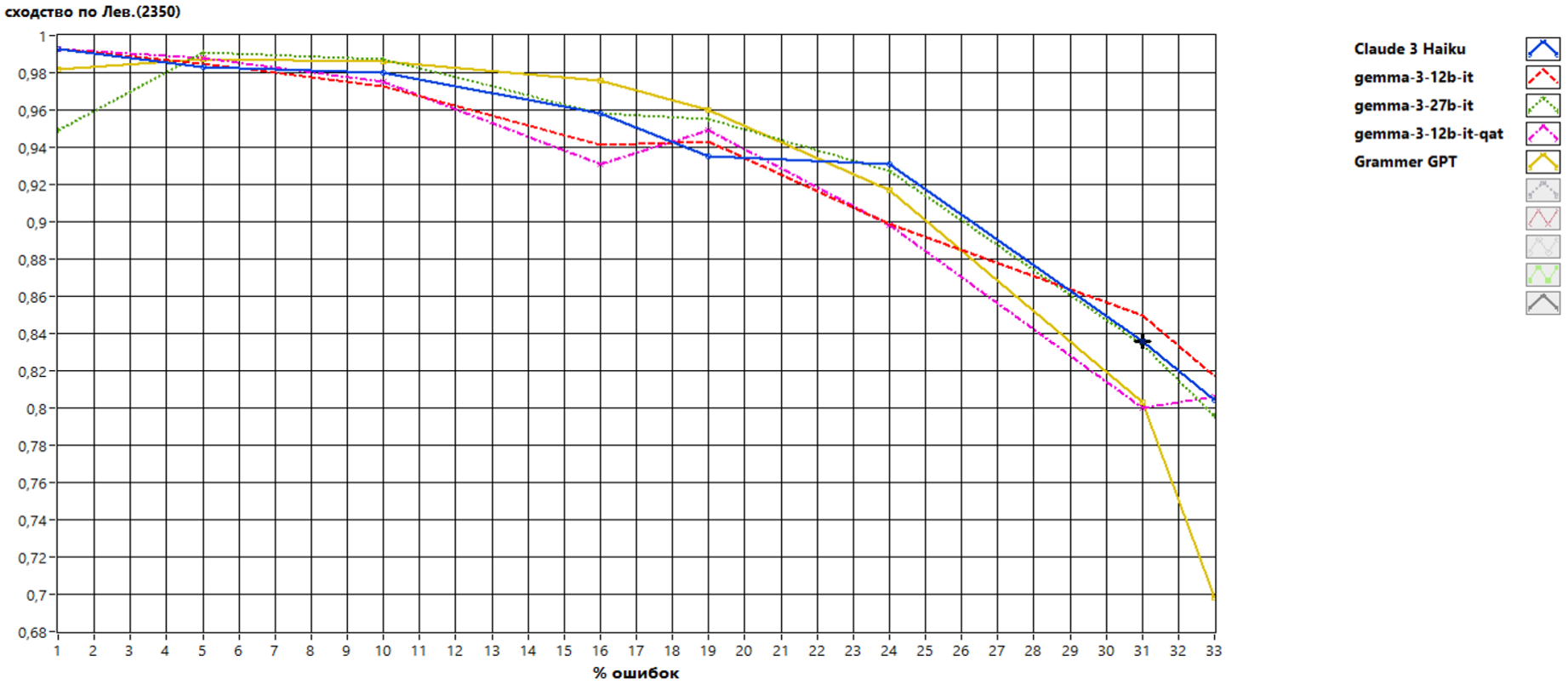
Зависимость скорости от уровня искажений

Точность коррекции зависит от размера offline моделей – 13B практически всегда лучше 7B (есть исключения в виде instruct моделей, дообучающихся на корпусе инструкций).
Для online моделей качество коррекции слабо зависит от объема (30B достаточно), но зависит от модели и оптимизации её параметров.
С ростом искажений точность коррекции падает.

Скорость коррекции текстов для разных моделей варьируется в диапазоне 50 до 250 симв/сек.
Зависит от размера модели, большие модели работают медленнее.
При сильных искажениях (более 24% текста) скорость резко уменьшается, и модели часто галлюцинируют.

ТОЧНОСТЬ (ЛЕВЕНШТЕЙН) КОРРЕКЦИИ В ЗАВИСИМОСТИ ОТ СТЕПЕНИ ИСКАЖЕНИЯ

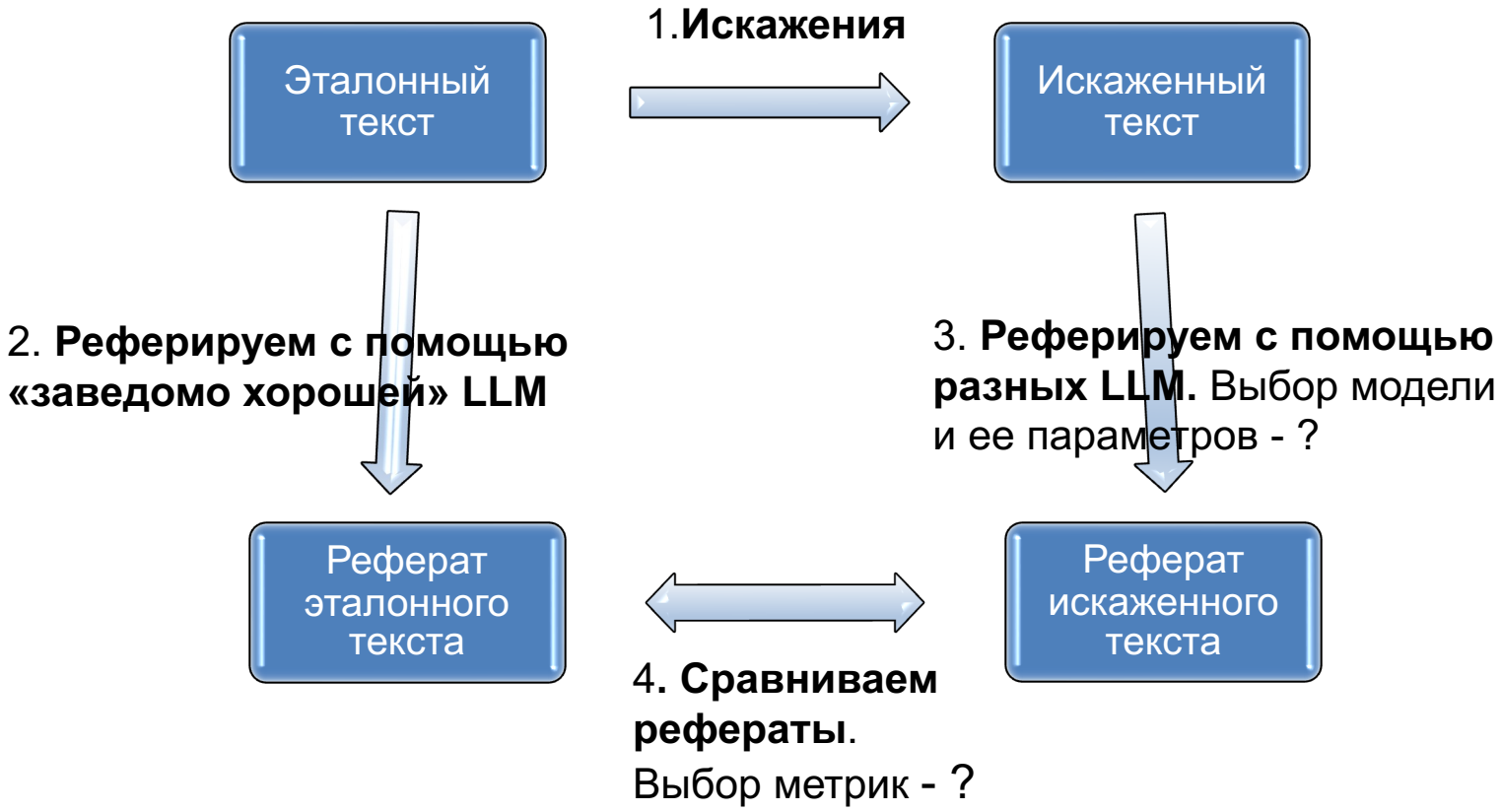
Точность коррекции (сходство по расстоянию Левенштейна между эталонным и скорректированным текстом) в зависимости от процента ошибок в искаженном тексте.



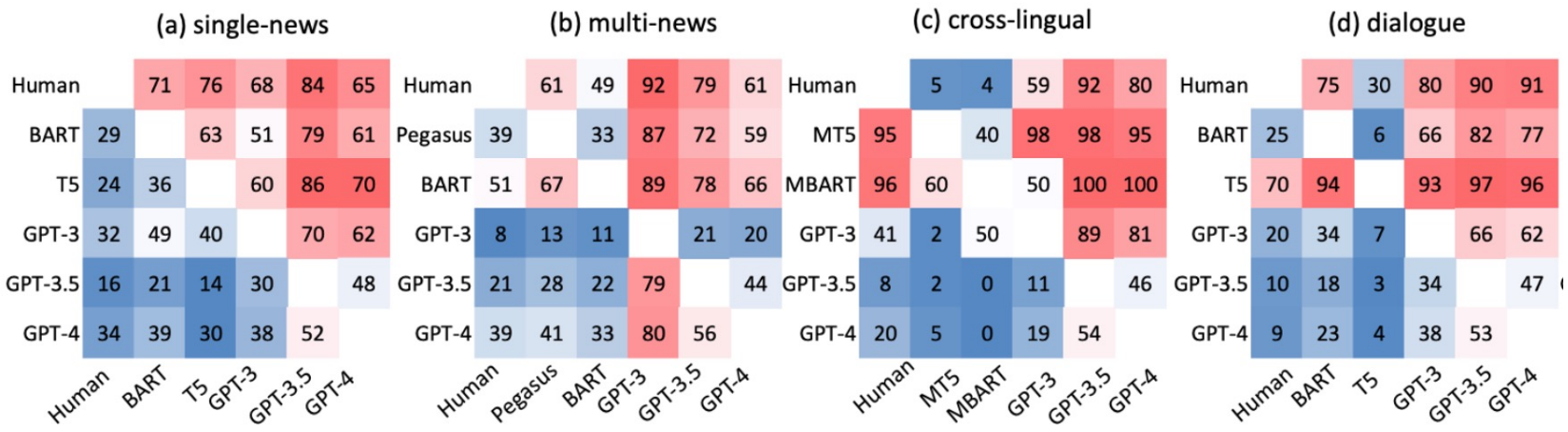
СУММАРИЗАЦИЯ. СХЕМА ИССЛЕДОВАНИЯ

Вопрос 1: Как искажения влияют на качество реферата?
Вопрос 2: Какие LLM лучше реферируют искаженные тексты?

Схема:



«SUMMARIZATION IS (ALMOST) DEAD»



Какие системы делают лучшие рефераты?

$(a)_{ij}$ – процент выигрыша рефератов, сделанными системой j, над рефератами сделанными системой i, по мнению эксперта – человека, на соответствующих корпусах.

Пример использования Prompt для модели GPT-3 (text-davinci-002 version)

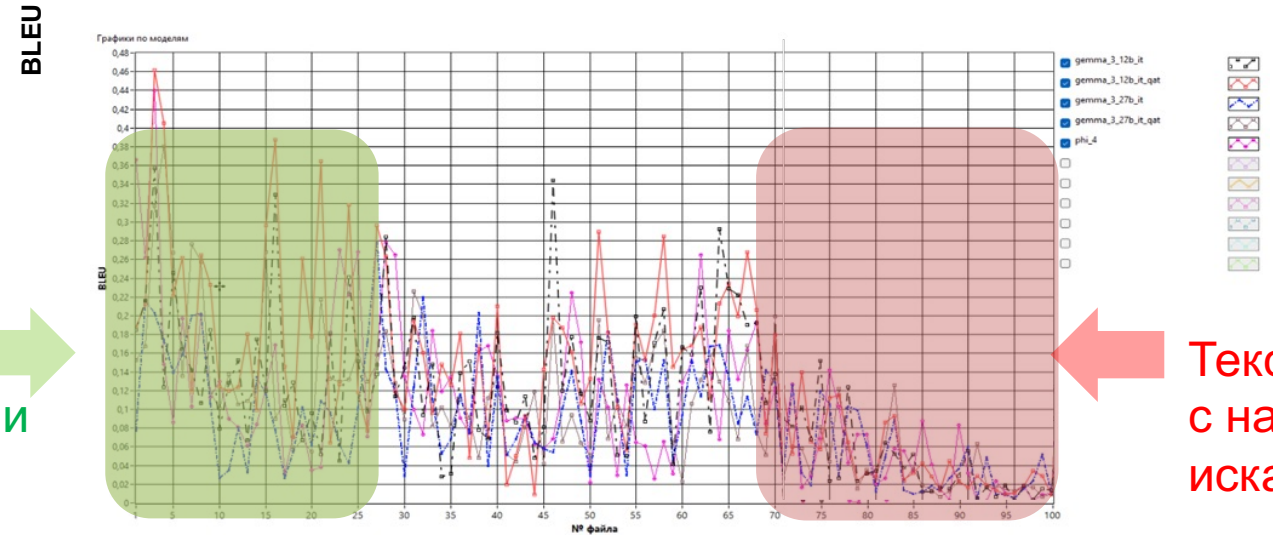
Goyal T., Li J. J., Durrett G. News summarization and evaluation in the era of gpt-3 //arXiv preprint arXiv:2209.12356. – 2022.
Pu X., Gao M., Wan X. Summarization is (almost) dead //arXiv preprint arXiv:2309.09558. – 2023.

РЕФЕРИРОВАНИЕ. РЕЗУЛЬТАТЫ

Исследовались метрики качества реферирования BLEU, ROUGE, METEOR.

Промпты:
1. "Пожалуйста, сделай краткое и точное резюме следующего текста на русском языке." (лучший в 189 тестах),
2. "Выдели основные идеи и ключевые моменты из этого текста. Сделай краткое содержание." (лучший в 87 тестах),
3. "Создай структурированное краткое содержание этого текста на русском языке, выделяя основные разделы и важные детали.",
4. "Напиши сжатое и понятное резюме этого текста на русском языке, не более 3-4 предложений.",
5. "Проанализируй этот текст и подготовь краткое аналитическое резюме, выделяя главные идеи, выводы и рекомендации." (лучший в 80 тестах).

Точность реферирования, 5 моделей, 100 текстов



Тексты с наименьшими искажениями

Тексты с наибольшими искажениями

ВЫВОДЫ

Основные результаты

1. Протестированы более 50 LLM с различными параметрами. По результатам тестов, проведенных летом 2025 года, лучшая offline модель – Meta-Llama-3-8b-instruct, лучшая online LLM – Claude 3 Haiku.
2. Online LLM модели лучше корректируют текст, чем offline модели, при искажениях < 30% количество ошибочных слов после коррекции уменьшается на 35-40%. Лучшие offline LLM модели эффективны при искажениях < 20%, количество слов с ошибками уменьшается на 25-30%.
3. Offline модели бОльшего объема обеспечивают в среднем лучшую точность коррекции.
4. При средних уровнях искажений качество реферата не сильно зависит от уровня ошибок.

Выявленные ограничения

1. Для устойчивой работы размер искаженного текста плюс запрос к модели (параметры модели + промпт) должны быть меньше максимального количества токенов для конкретной модели. Модель должна полностью загружаться в видеопамять.
2. При большом количестве ошибок скорость может упасть на порядок и модель может “подвиснуть”. Точность коррекции зависит от объема offline моделей и мало зависит от уровня квантования. Минимальный уровень квантования должен быть не меньше Q=4 и количество параметров при обучении не меньше 7B. Изменение параметров temperature, Top_p, и др. приводят к незначительным изменениям точности коррекции.
   Для online моделей точность коррекции слабо зависит от объема модели.

ИСПОЛЬЗОВАННЫЕ ИСТОЧНИКИ

1. Hajiali M. OCR Post-processing Using Large Language Models: дис. – University of Nevada, Las Vegas, 2023.
2. Thomas A., Gaizauskas R., Lu H. Leveraging LLMs for post-OCR correction of historical newspapers //Proceedings of the Third Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA)@ LREC-COLING-2024. – 2024. – С. 116-121.
3. Löfgren V., Dannélls D. Post-OCR correction of digitized Swedish newspapers with ByT5 //Proceedings of the 8th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL 2024). – 2024. – С. 237-242.
4. Jasonarson A. et al. Generating errors: OCR post-processing for Icelandic //Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa). – 2023. – С. 286-291.
5. de Araújo S. S. et al. A proposal for post-OCR spelling correction using Language Models //Latinx in AI@ NeurIPS 2024. – 2024.
6. Evaggelatos A. et al. Old Greek OCR Result Correction Using LLMs //Proceedings of the 2025 ACM Symposium on Document Engineering. – 2025. – С. 1-4.
7. Kanerva J. et al. OCR error post-correction with LLMs in historical documents: No free lunches //arXiv preprint arXiv:2502.01205. – 2025.
8. Boroş E. et al. post-correction of historical text transcripts with large language models: An exploratory study // Proceedings of the 8th joint SIGHUM workshop on computational linguistics for cultural heritage, social sciences, humanities and literature (LaTeCH-CLfL 2024). – 2024. – С. 133-159.
9. Biriuchinskii M., Alrahabi M., Roe G. Using LLMs for post-OCR correction on historical French texts: A case study using synthetic data //DH2025 CFP-Digital Humanities Conference 2025. – 2025.
10. Вершинин В. К., Ходненко И. В., Иванов С. В. Нормализация текста, распознанного при помощи технологии оптического распознавания символов, с использованием легковесных LLM //Электронные библиотеки. – 2025. – Т. 28. – №. 5. – С. 1036-1056.

По материалам исследования подготовлена статья:
Брицин А.А., Мельников С.Ю., Мещеряков Р.В., Пересыпкин В.А.
ОБ ЭФФЕКТИВНОСТИ ИСПОЛЬЗОВАНИЯ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ ДЛЯ КОРРЕКЦИИ ИСКАЖЕННЫХ ТЕКСТОВ

Работа частично поддержана грантом РНФ 24-11-00340
Исследование и разработка методов обработки слабоструктурированной информации на естественных языках в условиях сильных шумов для решения задач безопасности.

Российский научный фонд

СПАСИБО ЗА ВНИМАНИЕ

Спасибо за внимание!

Контакт:
Мельников Сергей Юрьевич
melnikov-syu@rudn.ru