

Историческая информатика

Правильная ссылка на статью:

Кузнецов А.В. Автоматическое извлечение информации из эго-документа: сравнительный анализ эффективности больших языковых моделей на примере дневника К.А. Березкина // Историческая информатика. 2025. № 3. С. 99-127. DOI: 10.7256/2585-7797.2025.3.75850 EDN: ZAYBBF URL: https://nbpublish.com/library_read_article.php?id=75850

Автоматическое извлечение информации из эго-документа: сравнительный анализ эффективности больших языковых моделей на примере дневника К.А. Березкина

Кузнецов Алексей Валерьевич

ORCID: 0000-0003-4755-250X

кандидат исторических наук

научный сотрудник; Институт всеобщей истории РАН

119334, Россия, г. Москва, Ленинский пр-т, 32 а, оф. 1426

✉ kuznetsovaleks@rambler.ru



[Статья из рубрики "Искусственный интеллект и наука о данных"](#)

DOI:

10.7256/2585-7797.2025.3.75850

EDN:

ZAYBBF

Дата направления статьи в редакцию:

11-09-2025

Дата публикации:

21-09-2025

Аннотация: Предметом исследования является сравнительный анализ производительности, аналитических стратегий и ограничений четырех больших языковых моделей – Gemini-2.5-Pro, o3, Grok3 и Deepseek-v3 – в задаче извлечения структурированной информации из исторического эго-документа. Анализ нацелен на определение способности моделей работать со сложным нарративом, характеризующимся высокой степенью субъективности, обилием косвенных свидетельств, многослойностью смыслов и эмоциональной окрашенностью. Ключевые

ограничения моделей – сверхинтерпретация, пропуск косвенных свидетельств и компромисс между полнотой и точностью – рассматриваются как часть их аналитических стратегий. Материалом послужил дневник вологодского гимназиста К.А. Березкина за 1849 год. Работа решает комплексную задачу по разработке и тестированию подхода, позволяющего трансформировать неструктурированный текст источника в датасет, пригодный для решения конкретной историографической задачи – анализа восприятия европейских революций 1848–1849 гг. в российской провинции. Методология основана на автоматическом извлечении структурированной информации при помощи больших языковых моделей. Разработан комплексный инструментарий, включающий доменно-специфическую онтологию, промпты и детальную JSON-схему для фиксации данных. Производительность моделей оценивалась на основе количественных (полнота, точность, F1-мера) и качественных показателей (гранулярность, точность следования онтологии, понимание исторического контекста, типичные ошибки). Научная новизна заключается в первой в отечественной историографии системной апробации и сравнительном анализе производительности ведущих языковых моделей при работе с историческим эго-документом. Установлено, что модели реализуют различные стратегии извлечения данных: от исчерпывающего, но «шумного» охвата (Gemini-2.5-Pro) до высокоточного, но избирательного (Deepseek-v3), что напрямую определяет пригодность получаемого датасета для разных исследовательских сценариев: от разведочного анализа до создания выверенных баз данных. Ключевой вывод исследования состоит в том, что автоматизированное извлечение является не технической операцией, а формой цифровой герменевтики. Соответственно, итоговый датасет – это не объективные данные (data), пассивно «обнаруженные» в источнике, а *corpora* – набор сведений, отобранных под конкретную задачу. Исследование показывает, что применение искусственного интеллекта повышает требования к критической экспертизе историка, смещая его роль от поиска информации к верификации и интерпретации машинных результатов.

Ключевые слова:

большие языковые модели, извлечение информации, искусственный интеллект, цифровая гуманитаристика, цифровая герменевтика, эго-документы, микроистория, революции 1848-1849 гг, Российская империя, XIX век

1. Введение

Появление больших языковых моделей не только расширяет инструментарий исторических исследований [6],[27], но и открывает новые перспективы для герменевтики исторических текстов [4],[26]. Особенно перспективным видится их применение к таким сложным для систематического изучения источникам личного происхождения, как дневники. Дневниковые записи, позволяют не только восстановить ход исторических событий, но и взглянуть на них глазами современников, проследить преломление глобальных процессов через призму индивидуального опыта. Но, сочетая в себе фактическую информацию, субъективное восприятие и имплицитные метаданные (время, место, упоминаемые лица и т.д.), они представляют собой особенно сложный материал для формализации. Именно эта многослойность, трудно поддающаяся традиционным методам анализа, делает их идеальным объектом для апробации новых цифровых подходов.

Настоящее исследование посвящено разработке методологии автоматического извлечения структурированной информации из дневниковых записей. В качестве

практического кейса для апробации этой методологии была выбрана историографическая задача анализа восприятия европейских революций 1848–1849 гг. в российской провинции. Реакция на революционные события 1848–1849 годов в Российской империи неоднократно становилась предметом изучения [9],[13], однако внимание историков традиционно концентрировалось на двух полюсах общественного мнения: резко негативной реакции консервативных кругов [8],[11] и сочувственном отношении оппозиционно настроенной интеллигенции [5],[12],[14],[15]. При этом восприятие революций в российских провинциях остается практически неизученным. Ценным источником для изучения этой темы может служить дневник вологодского гимназиста Кирилла Антоновича Березкина, который велся в 1849 году. Работа с подобным источником сопряжена со значительными методологическими сложностями: прямые упоминания о революционных событиях редки и вплетены в повседневную жизнь автора, тогда как текст содержит множество косвенных свидетельств (усиление цензуры, введение мер контроля в образовательных учреждениях, мобилизационные мероприятия и др.). Такая структура источника, где ключевая информация часто является неявной, делает его идеальным «полигоном» для проверки возможностей и ограничений больших языковых моделей. Их ценность в данном случае заключается не столько в скорости поиска очевидных фактов, сколько в способности к глубокому контекстуальному анализу: выявлению скрытых связей, различению фактологических, эмоциональных и идеологических слоев повествования, что представляет значительную сложность при традиционном «ручном» анализе большого массива текста.

Основная цель настоящего исследования – методологическая: разработать и апробировать подход к извлечению структурированной информации из сложного эго-документа с использованием больших языковых моделей. Для достижения этой цели решался ряд конкретных задач: с одной стороны, был создан программный инструментарий и разработана онтология для обработки дневника К. А. Березкина, с другой – проведен сравнительный анализ эффективности четырех флагманских моделей (Gemini-2.5-Pro, o3, Grok3 и Deepseek-V3) для формирования качественного набора данных (датасета). Выбор моделей обусловлен стремлением сопоставить производительность моделей последнего поколения (на момент проведения исследования в июне 2025 г.), представляющих ведущие конкурирующие архитектуры. На заключительном этапе был проведен количественный и качественный анализ полноты и точности извлечения с выявлением типичных ошибок каждой модели. Таким образом, ожидаемый вклад исследования выходит за рамки сугубо технического тестирования. Мы стремимся не только продемонстрировать применимость больших языковых моделей для цифровой обработки исторических источников и предложить практические ориентиры для выбора инструмента под конкретные исследовательские задачи, но и внести вклад в методологическую дискуссию о статусе автоматически извлеченных данных и роли историка в эпоху активного применения методов машинной интерпретации текста.

2. Обзор литературы

Извлечение информации (Information Extraction, IE) – это задача обработки естественного языка, направленная на автоматический поиск и структурирование фактической информации из неструктурированного текста. Ключевые подзадачи извлечения информации включают:

1) Распознавание именованных сущностей (Named Entity Recognition, NER): идентификация и классификация таких объектов, как имена людей, названия организаций, географические локации, даты и суммы.

2) Извлечение отношений (Relation Extraction, RE): обнаружение и классификация семантических связей между.

3) Извлечение событий (Event Extraction, EE): выявление событий и детальной информации о них: участников, места, времени и других атрибутов.

Ранние подходы к извлечению информации, основанные на наборе правил и статистических моделях, демонстрировали высокую эффективность в узкоспециализированных задачах, но с трудом масштабировались и испытывали трудности при работе с более сложными или разнородными данными, например, с историческими текстами [23],[33].

Возможности автоматической обработки исторических текстов существенно расширились с появлением больших языковых моделей. Они могут извлекать структурированную информацию даже без предварительного обучения на примерах (zero-shot) или на основе минимального количества примеров (few-shot). Это позволяет успешно решать задачи распознавания именованных сущностей, выявления связей между ними и извлечения событий [39].

Современные большие языковые модели общего назначения обучаются на колоссальных объемах разнородных данных, что обеспечивает им обширные «знания». Однако работа с историческими источниками требует учёта целого комплекса факторов: эволюции языка на протяжении веков, наличия устаревшей орфографии, возможных искажений при оптическом распознавании текста, а также необходимости интерпретировать сведения с учетом исторического контекста, избегая «галлюцинаций». Всё это значительно усложняет задачу извлечения информации [17],[18],[29].

Для преодоления ограничений языковых моделей исследователи используют два основных подхода: дополнительное обучение (fine-tuning) моделей на целевых данных и инжиниринг промптов (prompting), то есть управление поведением модели через тщательно составленные инструкции. Дополнительное обучение даёт хорошие результаты при адаптации модели к корпусу текстов определённого времени и жанра [37]. Однако этот подход сопряжён с серьёзными трудностями: необходимостью создания большого аннотированного обучающего корпуса [29],[37] и проблемой «катастрофического забывания», когда модель может утратить при дообучении часть своих обширных общих знаний [25]. Поэтому большинство исследований в области цифровой гуманитаристики опираются на инжиниринг промптов [22],[28],[34]. Ключевыми направлениями повышения точности в этом случае являются: разработка продуманных инструкций, обеспечивающих ясные критерии классификации, и предоставление дополнительной контекстуальной информации, отражающей исторический период и жанр источника [18],[25],[29].

Отдельное внимание в современной литературе уделяется интеграции доменно-ориентированных онтологий в процесс извлечения информации, что позволяет задать формализованную структуру категорий для классификации и минимизировать неоднозначность и «галлюцинации» модели. Данный подход находит применение в различных областях – от автоматического построения графов знаний [32] и иерархической классификации текста [39] до обработки исторических документов [38]. Использование онтологий оказывается особенно эффективным в задачах, требующих высокой структурированности и воспроизводимости результатов [20].

Особый интерес для нас представляют пока немногочисленные исследования, направленные на анализ личных нарративов, включая дневники. Недавние работы продемонстрировали потенциал больших языковых моделей для извлечения из эго-документов структурированной информации о целях их ведения, эпизодах саморефлексии и социальных связях [16],[21],[31],[35]. Эти работы демонстрируют возможность обработки субъективных и индивидуально окрашенных текстов.

Анализ литературы выявляет два пробела в существующих исследованиях. Во-первых, отсутствуют исследования, посвящённые комплексному извлечению разных типов структурированной информации – событий, общественных настроений и личных впечатлений – из исторических дневников с их уникальными содержательными и стилистическими особенностями. Во-вторых, не проводился систематический сравнительный анализ производительности ведущих больших языковых моделей на таком специфическом материале, который бы выявлял не только количественные метрики, но и качественные различия в их аналитических подходах. Настоящее исследование направлено на восполнение этих пробелов, предлагая методологию для анализа исторического эго-документа и всестороннюю оценку современных инструментов цифровой гуманитаристики.

3. Материал и метод исследования

3.1. Материал исследования

Дневник Кирилла Антоновича Березкина, ученика выпускного класса Вологодской губернской гимназии, охватывает период с 1 января по 12 сентября 1849 года и включает 249 записей. Содержание дневника отражает многогранный мир молодого человека середины XIX века: центральное место в нем занимают учебные будни в гимназии, светская жизнь провинциального дворянства, а также глубокая личная рефлексия автора о собственных поступках, чувствах и планах на будущее [7]. Полный электронный текст дневника доступен на странице проекта «Прожито» [2].

В рамках настоящего исследования текст дневника был преобразован в структурированный табличный формат (https://github.com/alexeyvkuznetsov/diary_event_extraction/blob/main/data/diary_with_id.csv). Таблица содержит 249 строк, каждая из которых соответствует одной дневниковой записи, и включает три столбца: `entry_id` (уникальный числовой идентификатор записи), `text` (полный текст записи) и `date` (дата записи в стандартизированном формате ГГГГ-ММ-ДД). Такая структура данных обеспечивает возможность последовательного анализа каждой записи с применением методов автоматического извлечения информации.

3.2. Разработка онтологии

Одним из ключевых элементов методологии стала разработка доменно-специфической онтологии (https://github.com/alexeyvkuznetsov/diary_event_extraction/blob/main/knowledge_map.json). Онтология выступила не просто классификационной схемой, но и формализованной моделью предметной области исследования – революций 1848–1849 гг., их европейского контекста, реакции Российской империи, а также личного восприятия этих событий автором дневника. Это позволило категоризировать релевантную информацию с высокой степенью детализации и обеспечивать единообразие результатов извлечения данных между разными моделями.

Онтология имеет древовидную структуру и представлена в формате JSON. Каждый элемент (узел) включает уникальный идентификатор (*id*), используемый для машинной классификации, и человекочитаемое наименование (*name*). Вершины верхнего уровня объединены в четыре блока, отражающих ключевые аспекты изучаемой проблематики: «Революционные события в Европе» (REV1848_EUROPE), «Реакция Российской Империи» (RU_REACTION_1848), «Идеологии и Причины» (IDEOLOGIES_CAUSES_1848) и «Личный опыт и восприятие автора» (AUTHOR_PERCEPTION_1848). Последний блок имеет особое методологическое значение, поскольку специально разработан для анализа эго-документов и позволяет четко разграничить собственное мнение и эмоциональную реакцию автора от информации о настроениях в обществе.

Тематика онтологии была намеренно задана шире, чем события, непосредственно отраженные в тексте дневника. Такой подход позволил не просто извлечь очевидные упоминания, но и оценить способность моделей отбирать релевантную информацию из более широкого семантического поля. Содержание онтологии итеративно дорабатывалось в ходе пробных попыток извлечения информации: вводились новые категории, чтобы избежать отнесения релевантных событий к неинформативной категории OTHER_1848 – «Другое / Не классифицировано».

3.3. Разработка инструментария

Для автоматизации процесса извлечения информации был разработан скрипт на языке Python

(https://github.com/alexeyvkuznetsov/diary_event_extraction/blob/main/src/info_extraction.py), осуществляющий пакетную обработку дневниковых записей, взаимодействие с большими языковыми моделями через API, формирование запросов и структурирование выходных данных.

Ключевым элементом методологии стала разработка промптов. Системный промпт определял роль модели как «высококвалифицированного историка-аналитика, специализирующегося на истории Европы и России XIX века» и задавал общую цель – извлечение информации, относящейся исключительно к революциям 1848–1849 гг. и их восприятию. Он также содержал необходимый исторический контекст об авторе источника, его социальном положении, хронологических и географических рамках событий, что позволяло модели интерпретировать текст как продукт определенной эпохи и социальной среды. Пользовательский промпт для каждой дневниковой записи объединял текст записи, ее дату, идентификатор, полный текст онтологии, а также детальные инструкции и примеры, иллюстрирующие ожидаемую структуру вывода и требуемую глубину анализа (Полностью промпты представлены в скрипте и в Приложении 1: https://github.com/alexeyvkuznetsov/diary_event_extraction/blob/main/doc/Supplement_1.pdf).

Для фиксации результатов была разработана детализированная JSON-схема, в которой фиксировались не только само событие и его классификация по онтологии, но и метаданные:

- *entry_id*: Уникальный идентификатор записи дневника.
- *event_id*: Идентификатор события из онтологии, позволяющий связать упоминание с типологией событий и реакций автора.
- *event_name*: Название события или аспекта из онтологии.

- *event_subtype_custom*: Дополнительное уточнение типа события для более детальной классификации.
- *description*: Подробное описание события на основе текста дневника с объяснением его связи с революциями 1848–1849 гг.
- *date_in_text*: Дата события, если она явно упомянута в тексте (отличается от даты записи).
- *source_date*: Дата дневниковой записи.
- *location*: Географическое место события, упомянутое в тексте.
- *location_normalized*: Нормализованное название места события на русском языке.
- *brief_context*: Краткий исторический контекст, содержащий один-два релевантных внешних исторических факта.
- *information_source*: Источник информации, на который ссылается автор дневника.
- *information_source_type*: Категория источника информации по типам: «Официальные источники (газеты, манифесты)», «Неофициальные сведения (слухи, разговоры в обществе)», «Личные наблюдения и опыт автора», «Информация от конкретного лица (именованный источник)», «Источник неясен/не указан».
- *confidence* («High»/«Medium»/«Low»): Оценка моделью уверенности в извлеченных данных.
- *classification_confidence* («High»/«Medium»/«Low»): Оценка моделью уверенности в правильности классификации события.
- *keywords*: Список ключевых слов и фраз для индексации и поиска.
- *text_fragment*: Точная цитата из дневника, содержащая релевантную информацию.

В качестве примера, иллюстрирующего структуру итогового датасета, ниже представлен JSON-объект для события, извлеченного из записи дневника от 7 января 1849 года моделью Gemini-2.5-Pro:

```
[
{
«entry_id»: 7,
«event_id»: «REV1848_HUN_MIL_AUS»,
«event_name»: «Военные действия (Австрия против Венгрии)»,
«event_subtype_custom»: «Успехи австрийских войск»,
«description»: «Автор, прочитав петербургские газеты, отмечает успехи австрийских войск в ходе подавления восстания в Венгрии.»,
«date_in_text»: null,
«source_date»: «1849-01-07»,
```

```

«location»: «Венгрия»,

«location_normalized»: «Венгрия»,

«brief_context»: «В декабре 1848 - январе 1849 гг. австрийская армия под
командованием фельдмаршала Виндишгреца вела успешное наступление против
венгерских революционных сил, заняв 5 января 1849 года столицу Венгрии Пешт.»,

«information_source»: «петербургские газеты»,

«information_source_type»: «Официальные источники (газеты, манифесты)»,

«confidence»: «High»,

«classification_confidence»: «High»,

«keywords»: [

«австрийцы»,

«Венгрия»,

«газеты»,

«военные действия»,

«успехи»

],

«text_fragment»: «Теперь читал петербургские газеты, иностранные происшествия. Везет
австрийцам, славно управляют в Венгрии, да что еще из этого всего будет?»

}

]

```

Данная структура позволяет не только извлечь фактическую информацию о событиях, но и зафиксировать способы получения информации автором, его субъективные оценки и эмоциональные реакции.

Полученный от языковой модели ответ в формате JSON проходил автоматическую валидацию для обеспечения соответствия заданной схеме данных (с использованием библиотеки Pydantic). Это гарантировало структурное единообразие всех итоговых файлов. Весь процесс сопровождался логированием и системой промежуточных сохранений, позволяющей возобновлять работу при технических сбоях.

3.4. Выбор и конфигурация больших языковых моделей

Для проведения сравнительного анализа были выбраны четыре флагманские (на момент работы над проектом) большие языковые модели, представляющие различные архитектуры и компании-разработчики:

1) Gemini-2.5-Pro – мультимодальная модель от Google (*Компания-владелец Google признана в Российской Федерации экстремистской организацией*). Отличается большим контекстным окном (до 1 миллиона токенов) и демонстрирует передовые возможности в решении задач, требующих логического мышления;

2) o3 – мультимодальная модель от OpenAI, разработанная с акцентом на улучшенные способности к рассуждению. o3 использует технику, позволяющую ей «обдумывать» ответ, выполняя промежуточные шаги для решения сложных задач. Эта модель демонстрирует превосходные результаты в решении сложных научных и технических задач;

3) Grok-3 – мультимодальная модель от xAI, которая использует архитектуру смеси экспертов (Mixture of Experts, MoE) с примерно 2,7 триллионами параметров. Она разработана для решения сложных задач и имеет контекстное окно в 128 000 токенов. Модель отличается менее строгой «цензурой» в ответах, в отличие от других коммерческих моделей;

4) Deepseek-V3 – модель с открытым исходным кодом от компании DeepSeek. Она построена на архитектуре смеси экспертов, имеет 671 миллиард общих параметров, из которых для каждого запроса активируется только 37 миллиардов, что делает её исключительно эффективной с точки зрения стоимости. Модель имеет контекстное окно в 128 000 токенов.

Все запросы к моделям выполнялись с одинаковыми настройками, значение параметра temperature выставлялось равным 0.1 для снижения случайности и повышения воспроизводимости результатов, что является критически важным для задач извлечения структурированной информации.

3.5. Процедура извлечения и методика оценки

Процесс автоматизированного извлечения информации был последовательно выполнен для четырех выбранных моделей. Каждая модель обработала все 249 записей дневника, генерируя на выходе структурированные данные, которые сохранялись в отдельные JSON-файлы: `deepseek-v3.json`, `gemini-2.5-pro.json`, `grok3.json`, `o3.json` (https://github.com/alexeyvkuznetsov/diary_event_extraction/tree/main/results/model_outputs). Все исследуемые языковые модели демонстрировали устойчивое следование инструкциям промпта и формировали ответы, соответствующие заданной JSON-схеме. Лишь в единичных случаях требовались автоматические корректировки формата данных для приведения их к единой схеме.

Для оценки производительности моделей была проведена ручная разметка текста дневника, на основе которой был создан эталонный набор данных («золотой стандарт»), включающий 34 дневниковые записи, содержащие релевантную теме исследования информацию

(https://github.com/alexeyvkuznetsov/diary_event_extraction/blob/main/data/gold.csv).

Важно подчеркнуть, что «золотой стандарт» создавался исключительно для верификации и сопоставления результатов и не использовался в процессе извлечения. Каждая из моделей самостоятельно обрабатывала все 249 записей дневника, чтобы выявить из них значимые фрагменты.

Оценка производилась путем сопоставления результатов каждой модели с «золотым стандартом» в рамках сводной таблицы (См. Приложение 1: https://github.com/alexeyvkuznetsov/diary_event_extraction/blob/main/doc/Supplement_2.pdf), в которой для каждой записи из унифицированного списка (включающего все упоминания из «золотого стандарта» и результатов моделей) фиксировался статус извлечения:

1) Истинно-положительное срабатывание (True Positive, TP): Модель корректно идентифицировала релевантную запись, присутствующую в «золотом стандарте».

2) Ложно-положительное срабатывание (False Positive, FP): Модель ошибочно идентифицировала запись как релевантную (ошибка первого рода).

3) Ложно-отрицательное срабатывание (False Negative, FN): Модель пропустила релевантную запись, присутствующую в «золотом стандарте» (ошибка второго рода).

На основе этих данных были рассчитаны три стандартные метрики из области извлечения информации:

1) Полнота (Recall) = $TP / (TP + FN)$, отражающая долю всех релевантных событий, которую смогла обнаружить модель.

2) Точность (Precision) = $TP / (TP + FP)$, показывающая долю релевантных событий среди всех извлеченных моделью.

3) F1-мера (F1-score) = $2TP / (2TP + FP + FN)$, представляющая собой гармоническое среднее полноты и точности и служащая сбалансированным показателем общей производительности.

Наряду с количественными измерениями, был проведен углубленный качественный анализ, направленный на изучение характера извлеченных данных и допущенных ошибок, включая оценку гранулярности извлечения, качества генерируемых описаний и корректности классификации по типам свидетельств (прямые и косвенные).

4. Результаты

4.1. Количественный сравнительный анализ

Сравнительный анализ количественных показателей продемонстрировал фундаментальные различия в стратегиях извлечения информации, применяемых каждой из тестируемых моделей. Полученные данные (Таблица 1) свидетельствуют о классической для задач извлечения информации обратной зависимости между полнотой и точностью: повышение одного показателя, как правило, ведет к снижению другого.

Таблица 1. Сводные метрики производительности моделей

Модель	TP	FP	FN	Полнота (Recall)	Точность (Precision)	F1-Score
Gemini-2.5-Pro	34	21	0	100%	61,8%	76,4%
o3	26	2	8	76,5%	92,9%	83,9%
Grok3	30	5	4	88,2%	85,7%	86,9%
Deepseek-V3	27	1	7	79,4%	96,4%	87,1%

Модель Gemini-2.5-Pro показала стратегию максимального охвата, достигнув стопроцентной полноты (100%) и успешно идентифицировав все 34 релевантные записи. Это гарантирует, что исследователь не пропустит ни одного потенциально важного фрагмента. Однако такой исчерпывающий подход достигается ценой существенного снижения точности (61,8%), что выразилось в 21 ложном срабатывании.

В противоположность этому, остальные модели демонстрируют значительно более высокую точность. Лидером по этому параметру стала Deepseek-V3 с почти безупречной точностью в 96,4% и всего одним ложным срабатыванием. Однако эта надежность достигается ценой снижения полноты (79,4%): модель пропустила 7 релевантных записей. Схожую производительность показала o3 с точностью 92,9% и всего двумя ложными срабатываниями, но при этом с 8 пропущенными записями. Модель Grok3

занимает наиболее сбалансированную позицию с высокими показателями как полноты (88,2%), так и точности (85,7%).

Таким образом, уже на количественном уровне отчетливо выделяются два принципиально разных подхода к извлечению данных: исчерпывающий, но «шумный» (Gemini-2.5-Pro), и точный, но более избирательный (Deepseek-V3, o3). Это подчеркивает, что выбор модели является не техническим, а методологическим решением, определяющим саму стратегию исследования: либо первоначальный максимально широкий поиск для формирования гипотез, либо точечное извлечение данных по уже известным параметрам.

4.2. Анализ извлечения по типам данных

Выявленные различия в стратегиях становятся особенно наглядными при анализе способности моделей работать с разными типами свидетельств. Для исторического исследования критическое значение имеет способность различать не только прямые фактологические упоминания, но и косвенные. Наш эталонный набор включает 20 записей с прямыми упоминаниями событий и 14 записей с косвенными свидетельствами.

Таблица 2. Полнота извлечения по типам свидетельств (прямые – косвенные)

Модель	Прямые	Косвенные
Gemini-2.5-Pro	100% (20/20)	100% (14/14)
o3	100% (20/20)	42,9% (6/14)
Grok3	100% (20/20)	71,4% (10/14)
Deepseek-v3	100% (20/20)	50% (7/14)

Как показывают данные, представленные в Таблице 2, извлечение прямых упоминаний (новости из газет, публикация манифеста) не вызвало затруднений ни у одной из моделей: все они показали 100% полноту по этому типу данных. Однако ключевые расхождения проявились при работе с косвенными свидетельствами. Именно здесь Gemini-2.5-Pro показала себя как наиболее чувствительный инструмент, способный улавливать скрытые в бытовых описаниях маркеры эпохи.

Наглядным примером является запись от 8 августа 1849 года, где автор упоминает, что полк его знакомого «стоит теперь вместо гвардии в Петербурге». Только Gemini-2.5-Pro, опираясь на фоновые знания, смогла корректно интерпретировать эту фразу как косвенное свидетельство мобилизации, в ходе которой элитные гвардейские части были переброшены из столицы ближе к театру военных действий. Другие модели пропустили этот фрагмент, поскольку он не содержит явных ключевых слов, что демонстрирует их ограниченную способность к глубокой контекстуальной интерпретации скрытых смыслов, которая и отличает работу профессионального историка. Именно в этом заключается ключевое преимущество использования наиболее совершенных языковых моделей: они способны выступать не просто как поисковый инструмент, а как «ассистент-аналитик», обращающий внимание исследователя на неявные связи, которые могли бы быть упущены при традиционном «ручном» чтении, ориентированном на поиск прямых упоминаний.

4.3. Сравнительный анализ гранулярности извлечения и точности следования онтологии

Эффективность больших языковых моделей в задачах извлечения структурированной информации из исторических источников определяется не только способностью

обнаруживать релевантные фрагменты, но и двумя взаимосвязанными параметрами: гранулярностью извлечения и точностью следования предложенной онтологии. Гранулярность характеризует способность модели декомпозировать многослойный нарратив на отдельные семантические единицы (факт, источник информации, авторская оценка), а точность следования онтологии – умение модели корректно классифицировать извлеченные единицы и обогащать их осмысленными метаданными. Именно сочетание этих качеств отличает поверхностное извлечение фактов от глубокого аналитического прочтения источника. Продемонстрируем различия в аналитических способностях моделей на трех показательных записях из дневника.

Текстовый фрагмент из записи от 7 января 1849 г. представляет собой яркий пример сложности источника личного происхождения, где в коротких фразах совмещаются фактология, мнение, эмоция и рефлексивный вопрос о будущем: *«Теперь читал петербургские газеты, иностранные происшествия. Везет австрийцам, славно управляют в Венгрии, да что еще из этого всего будет? Ералаш ужаснейший, кажется, еще и теперь на Западе»*. Фрагмент содержит как минимум четыре различных семантических слоя: (1) констатацию факта из газет, (2) выражение личного одобрения действий австрийской монархии, (3) рефлексию о неопределенности будущего и (4) обобщенную негативную оценку общеевропейской ситуации. Успешность анализа этого фрагмента зависит от способности языковой модели не просто извлечь факты, а распутать сложный клубок авторского сознания.

Gemini-2.5-Pro демонстрирует наиболее глубокое прочтение текста. Модель создала четыре записи полностью раскрыв структуру текстового фрагмента. Она безошибочно отделила полученную из газет информацию о военных успехах Австрии в Венгрии (событие «Военные действия (Австрия против Венгрии)») от личного, идеологически окрашенного одобрения этих действий («Выражение поддержки/осуждения действий властей/конкретных государств»). Далее, модель так же успешно разделила обобщенную констатацию хаоса на континенте («Общие упоминания Весны народов, беспорядков на Западе») и персональную тревогу автора о последствиях этих событий, выраженную в риторическом вопросе («Размышления о политике, войне, обществе, будущем»). Такая степень детализации позволяет исследователю работать с данными на микроуровне, анализируя не только то, что знал автор, но и как он это осмыслял и переживал.

Высокую гранулярность, сопоставимую с лидером, показали модели o3 и Grok3, однако их интерпретации выявляют специфические аналитические акценты. Модель o3 также успешно создала четыре записи, корректно разграничив факт и мнение. Ее классификация авторского восприятия, однако, была более общей: она определила реакцию на «ералаш ужаснейший» как «Общая эмоциональная реакция/сильное впечатление», что верно, но менее конкретно, чем у Gemini-2.5-Pro. Модель Grok3, в свою очередь, предложила интересную интерпретацию, сфокусированную на эмоциональной составляющей. Она классифицировала авторскую реакцию не как политическую поддержку или рефлексию, а как проявление конкретных эмоций из онтологии: риторический вопрос о будущем был отнесен к категории «Любопытство, интерес», а оценка хаоса на Западе – к категории «Страх, беспокойство, тревога». Такой подход, хотя и упускает политический аспект оценки, может быть ценным для исследований в области истории эмоций, так как модель демонстрирует способность распознавать аффективные сигналы в тексте.

На этом фоне результат работы DeepSeek-V3 предстает как существенное упрощение. Модель создала всего две записи. В первом она объединила факт (успехи австрийцев) и личное мнение автора о них, тем самым стерев ключевую для историка грань между

информацией и ее рецепцией. Во втором событии аналогичным образом были объединены общая характеристика европейских событий и личная тревога автора. В результате такой редукции утрачивается важнейший слой информации, касающийся субъективного восприятия, что делает извлеченные данные неполными для глубокого исторического анализа.

Запись от 6 апреля 1849 г. ставит перед моделями ещё более сложную задачу. Текст начинается с фиксации циркулирующих в обществе слухов о возможной войне с Турцией, затем переходит к упоминанию новостей из иностранной прессы и завершается развернутой личной рефлексией автора о собственном изменившемся отношении к войне, которая, в свою очередь, распадается на самоанализ и романтические фантазии о героической смерти. Эта многослойность стала настоящим испытанием для моделей.

Наиболее высокую аналитическую производительность вновь продемонстрировала Gemini-2.5-Pro, которая смогла декомпозировала нарратив на четыре семантически завершенных части, точно отражающих его внутреннюю структуру. Во-первых, модель корректно идентифицировала общественный дискурс, классифицировав его как «Обсуждение событий / Слухи в обществе». Во-вторых, она выделила в отдельную запись упоминание сведений из французской газеты, отнеся его к категории «Распространение новостей/слухов о революциях», что точно фиксирует канал поступления информации. В-третьих, модель разделила личную рефлексию автора на два уровня: общие размышления об изменившемся отношении к военной службе («Размышления о политике, войне, обществе, будущем») и конкретное проявление этого в воображаемой ситуации – романтические мечты о поступлении на службу из-за несчастной любви («Прямое влияние событий на жизнь/планы автора»). Последняя запись, снабжена подтипом «Война как фон для романтических фантазий», что свидетельствует о незаурядной способности модели к интерпретации сложных культурных и психологических сюжетов.

Прочие модели, напротив, продемонстрировали разную степень редукции смыслов исходного текста. DeepSeek-V3, как и в предыдущем случае, показала низкую гранулярность, объединив все слухи и новости в одну категорию («Обсуждение событий / Слухи в обществе»), а всю многогранную рефлексию автора – в другую («Размышления о политике, войне, обществе, будущем»). Такой подход, хотя и верен по сути, теряет важные детали. Модель o3 прибегает к еще более существенной редукции, сводя сложный процесс самоанализа автора к единственной эмоции «Страх, беспокойство, тревога», что является сильным упрощением. Кроме того, она неверно классифицирует слух об отправке посланника в Константинополь как свершившийся факт («Дипломатические действия / Заявления»), что размывает границу между слухом и реальностью.

Наименее релевантные результаты показала модель Grok3, которая полностью проигнорировала обширный пласт текста, посвященный личной рефлексии. Модель извлекла только информацию о слухах, классифицировав их как «Дипломатические действия / Заявления», и «Обсуждение событий / Слухи в обществе».

Наконец, запись от 5 мая 1849 года представляет собой настоящий «стресс-тест» для аналитических возможностей моделей, требуя от них проанализировать развернутый нарратив, в котором сплетаются слухи настоящего времени, официальные новости, личные аналитические выводы, вспышка ксенофобской ненависти и ретроспективное обращение к слухам и страхам прошлого года. Этот текст является сложнейшим для формализации каскадом различных информационных и психологических состояний

автора: от фиксации мобилизационных слухов («еще вчера говорили, для чего бы это велено собрать всех бессрочно отпущенных солдат») к официальной новости из газеты (публикация «Манифеста о движении армий наших для содействия Императору Австрийскому на потушение мятежа в Венгрии и Трансильвании»), далее к собственному умозаключению (поляки как главные виновники мятежа), которое перерастает в эмоциональный взрыв («перевешать, как скотов»), и, наконец, к воспоминанию о страхах осени 1848 года, подкрепленному пересказом слухов о заговоре поляков с целью поджогов российских городов.

В этом сложном задании Gemini-2.5-Pro вновь продемонстрировала высокий уровень анализа, разделив текст на семь логически завершенных и точно классифицированных событий. Модель безошибочно разграничила слухи, классифицированные как «Мобилизация / Передвижение войск» (с подтипом «Сбор бессрочно отпущенных солдат и слухи о войне»), от официальной новости о «Манифестах о вмешательстве / войне». Аналогичным образом, она отделила фактологическую информацию из Манифеста о «Польских участниках в европейских революциях» от основанного на ней авторского вывода, отнесенного к категории «Личные предубеждения/симпатии/антипатии автора». Наконец, когнитивный акт предубеждения был четко отделен от аффективного взрыва, выделенного в самостоятельную категорию «Гнев, возмущение, ненависть».

Особенно показательна работа модели с ретроспективной частью текста. Gemini-2.5-Pro создала две отдельные записи: одну для фиксации самого слуха о заговоре поджигателей, отнесенное к категории «Обсуждение событий / Слухи в обществе», и вторую для описания общей атмосферы того времени, классифицированную как «Страхи / Опасения перед революцией/беспорядками». Более того, модель эффективно использовала поле *event_subtype_custom*, снабдив каждую запись емким описанием, что многократно повышает ценность извлеченных данных.

Модели o3 и Grok3 также продемонстрировали высокую гранулярность, извлекая восемь и семь событий соответственно, и в целом справились с задачей, но с характерными недостатками. o3, как и в предыдущем случае, показала склонность к еще более дробному психологическому анализу, разделив личную реакцию автора на три категории: «Личные предубеждения/симпатии/антипатии автора», «Гнев, возмущение, ненависть» и «Размышления о политике, войне, обществе, будущем». Однако ее ключевой недостаток – полное игнорирование поля *event_subtype_custom*, что делает ее результаты менее информативными. Grok3, в свою очередь, верно идентифицировала большинство событий, но при анализе воспоминаний автора объединила слух и страх в одну категорию «Страхи / Опасения перед революцией/беспорядками», что является некоторым упрощением по сравнению с Gemini-2.5-Pro.

DeerSeek-V3 вновь показала себя как аутсайдер в плане глубины анализа. Модель объединила слухи о сборе солдат с уже идущим в обществе обсуждением начала войны в одну категорию «Обсуждение событий / Слухи в обществе», упустив важный для понимания динамики распространения информации переход от вопроса к утверждению. Критичнее всего для исследователя оказалось то, как модель интерпретировала эмоционально напряженную часть текста. Фраза «перевешать, как скотов», являющаяся маркером языка вражды, не была идентифицирована как проявление ненависти, а оказалась нивелирована в общей категории «Личные предубеждения/симпатии/антипатии автора». Модель оказалась неспособна отличить идеологическое предубеждение от прямого вербального выражения крайней степени агрессии.

Качественный анализ подтверждает наличие иерархии в аналитических возможностях моделей. Gemini-2.5-Pro последовательно демонстрирует себя как тонкий аналитик, способный различать семантические слои, критически важные для исторического исследования. Остальные модели, напротив, проявляют склонность к различным формам редукционизма: от агрегации данных и упрощения (DeepSeek-V3) до узкой специализации на отдельных аспектах, таких как эмоции (Grok3), что ведет к потере многомерности эго-документа.

4.4. Понимание исторического контекста

Еще один существенный аспект, определяющий ценность большой языковой модели как инструмента для исторических исследований, – её способность понимать исторический контекст. В нашем случае это проявляется в качестве заполнения полей *description* (описание события) и *brief_context* (краткая историческая справка).

Заполнение поля *brief_context* показывает способность модели обогащать извлеченные данные релевантной внешней информацией, помещая запись из дневника в корректную событийную рамку. Наилучшие результаты в этом продемонстрировала модель Gemini-2.5-Pro, выступая в роли компетентного ассистента-историка. Например, при обработке записи от 7 января 1849 г., где автор упоминает успехи австрийцев в Венгрии, она предоставила высокодетализированный контекст: «В декабре 1848 - январе 1849 гг. австрийская армия под командованием фельдмаршала Виндишгреца вела успешное наступление против венгерских революционных сил, заняв 5 января 1849 года столицу Венгрии Пешт». Эта информация не только точна, но и хорошо объясняет новостной фон, на который реагировал автор дневника.

В отличие от этого модели o3 и Grok3 предоставили корректный, но значительно более общий контекст. Для той же записи их формулировки, такие как «Зимой 1848–1849 гг. Австрия начала контрнаступление и заняла ряд венгерских городов» (o3) и «В 1848–1849 гг. Австрийская империя вела активные военные действия для подавления венгерского восстания, добившись значительных успехов к началу 1849 года» (Grok3), являются верными, но лишены конкретики. Модели успешно определяют общий тренд событий, но не связывают его с точным моментом времени, что снижает ценность справки. Самый низкий уровень исторической контекстуализации показала DeepSeek-V3, которая в большинстве случаев ограничивалась перефразированием уже имеющейся в исходном тексте информации. Ее описание контекста к записи от 7 января 1849 г. – «В 1848–1849 гг. Австрия вела военные действия против венгерских революционеров, стремясь подавить восстание» – фактически дублирует описание события.

Эта же иерархия производительности наблюдается и при заполнении поля *description*, отражающего способность модели к аналитическому синтезу информации из самого текстового фрагмента. Gemini-2.5-Pro вновь демонстрирует лучшие результаты, поскольку ее описания носят характер не простого пересказа, а микроанализа. Наиболее показательным является обработка записи от 1 марта 1849 г., где автор дневника пишет о вводе русских войск в Трансильванию «на помощь австрийцам против чехов». Gemini-2.5-Pro не только корректно извлекает событие, но и добавляет в описание ключевой комментарий: «Он ошибочно называет противников «чехами», хотя речь идет о венгерских революционных силах...». Эта способность к верификации информации из источника превращает описание в полноценный научный комментарий.

Модель o3 также генерирует качественные и точные описания, которые ясно и недвусмысленно передают суть события. Описания моделей Grok3 и DeepSeek-V3

зачастую выглядят как упрощенное перефразирование исходного текста без попытки его аналитического осмысления.

Вместе с тем, именно эти поля, демонстрирующие «интеллект» модели, требуют от исследователя наибольшего внимания. Ни одна модель не застрахована от фактических ошибок и «галлюцинаций». Например, в рассмотренной нами выше записи от 6 апреля 1849 г. слухи о возможной войне с Турцией Gemini-2.5-Pro в поле *brief_context* комментирует следующим образом: «В 1849 году отказ Османской империи выдать России и Австрии венгерских и польских революционеров (включая Лайоша Кошута), бежавших после подавления восстания, привел к острому дипломатическому кризису и угрозе войны». Такой дипломатический кризис действительно имел место, но произошел осенью 1849 г., после подавления восстания. [1, с. 221-265]. Напряженность же весной 1849 г. была связана с неопределённостью вокруг статуса Дунайских княжеств и закончилась заключением компромиссной Балта-Лиманской конвенции (19 апреля (1 мая) 1849 г.), которая закрепила совместный контроль России и Османской империи над Дунайскими княжествами [10, с. 223-234]. Данный пример служит важнейшим методологическим предостережением: даже самая продвинутая модель может генерировать правдоподобные, но фактически неточные контекстуальные справки. Это в очередной раз доказывает, что автоматизированное извлечение данных не отменяет, а, напротив, повышает требования к критической экспертизе и эрудиции самого историка, который остается финальным арбитром в интерпретации источника и его контекста. Работа с языковой моделью выстраивается не как делегирование полномочий, а как диалог, в котором модель предлагает структурированную гипотезу и первоначальный контекст, а историк производит их критическую верификацию и окончательную научную интерпретацию.

4.5. Сравнительный анализ ошибок

Анализ допущенных ошибок – ложных срабатываний и пропусков – позволяет заглянуть в «черный ящик» и понять «слепые зоны» каждой модели. Это не просто подсчет неточностей, а качественная диагностика их аналитических подходов.

Ложные срабатывания – это случаи, когда модель извлекала информацию из записи, которую эксперт счел нерелевантной. Анализ позволил выделить три основные категории таких ошибок. Анализ позволил выделить три основные категории таких ошибок, количественное распределение которых по моделям показано в Таблице 3:

- 1) Концептуальная сверхинтерпретация: модель делает логически правдоподобный, но текстуально не подтвержденный вывод, проецируя свои макроисторические знания на локальное событие.
- 2) Неверное определение релевантности: модель извлекает фрагменты, описывающие личную жизнь автора (чтение, карьерные переживания), и ошибочно присваивает им статус общественно-политического события.
- 3) Извлечение «мета-информации»: специфическая ошибка Gemini-2.5-Pro, когда сообщение об отсутствии информации (например, «в газетах нет ничего примечательного») фиксируется как самостоятельное событие.

Таблица 3. Типология ложных срабатываний (FP) по моделям

Тип ошибки	Gemini-2.5-Pro (21)	Grok3 (5)	o3 (2)	Deepseek-v3 (1)

Концептуальная сверхинтерпретация	11	4	2	1
Неверное определение релевантности	8	1	-	-
Извлечение «мета- информации»	2	-	-	-

Наиболее частые ошибки моделей – именно сверхинтерпретация. Лидер по общему количеству ошибок – Gemini-2.5-Pro. В большинстве случаев исходя из своих представлений о «мрачном семилетии», о чем модель пишет в поле *brief_context*, модель проецирует этот макро-контекст на микро-события из жизни автора (меры по поддержанию дисциплины в гимназии, личные переживания), не имея для этого достаточных оснований в самом тексте. Например, в записи от 22 августа 1849 года автор описывает подозрительное поведение ученика и заключает: «дышит изменой и подлостью в гимназии». Gemini-2.5-Pro связала это с общей атмосферой доносившего в «мрачное семилетие». Ошибки неверного определения релевантности связаны с реакцией автора на чтение литературных произведений Эжена Сю. Важно отметить, что в 7 из 21 случая ложных срабатываний Gemini-2.5-Pro самостоятельно понижала оценку в поле *confidence* до значения «Medium», что может служить полезным флагом для исследователя при последующей верификации.

Если ложные срабатывания создают «шум», который можно отфильтровать, то пропуски ведут к прямой потере исторической информации и являются более серьезной проблемой для исследователя. В то время как Gemini-2.5-Pro не допустила ни одного пропуска релевантной информации, другие модели продемонстрировали значительные упущения: Grok3 – 4, o3 – 8, а Deepseek-V3 – 7.

Принципиально важно, что все без исключения пропущенные записи относились к категории косвенных свидетельств. Этот вывод полностью подтверждает результаты, полученные в разделе 4.2, и демонстрирует, что главный риск при использовании моделей с высокой точностью (o3, Deepseek-V3) заключается в их неспособности улавливать нюансы, намеки и скрытые смыслы, которые и составляют особую ценность эго-документов.

4.6. От текста к данным: практическая апробация результатов

Полученный на основе ответов моделей структурированный датасет в формате JSON является готовой основой для количественного анализа – например, для выявления статистических закономерностей в динамике внимания автора к событиям или в частоте его обращения к разным типам источников. Пример такого разведочного анализа представлен в Jupyter-ноутбуке, доступном в репозитории проекта (https://github.com/alexeyvkuznetsov/diary_event_extraction/blob/main/diary_analysis.ipynb). Чтобы наглядно продемонстрировать практическую ценность предложенной методологии и показать, как именно трансформация нарративного текста в структурированные данные обогащает исследовательский процесс, полученный на основе ответов моделей датасет был использован для создания интерактивного дашборда (https://alexeyvkuznetsov.github.io/diary_dashboard/). Этот инструмент является не просто визуализацией, а доказательством состоятельности предложенной методологии, наглядно показывая, как она трансформирует нарративный источник в многомерную аналитическую базу данных. За основу был взят наиболее полный результат модели Gemini-2.5-Pro, из которого после ручной верификации были удалены ложные срабатывания (Финальный датасет см. по ссылке:

https://github.com/alexeyvkuznetsov/diary_event_extraction/blob/main/revolution_events.json).

Ключевым элементом дашборда является система динамической фильтрации, позволяющая исследователю производить целевой анализ, сегментируя корпус извлеченных данных по множеству параметров: временному периоду, категории из онтологии, географии, типу источника информации. Это иллюстрирует переход от традиционного «медленного чтения» источника, через этап автоматизированного извлечения информации, к ее интерактивному анализу: можно в несколько кликов отследить географию событий (от локальных в Вологде, до глобальных), выявить доминирующие темы в определенный месяц или проанализировать, из каких источников (газеты или слухи) автор черпал информацию о конкретных событиях.

Таким образом, дашборд – это полноценная исследовательская среда, доказывающая, как применение больших языковых моделей для структурирования исторических эго-документов открывает новые возможности для анализа и проверки гипотез, переходя от традиционного чтения к активному взаимодействию с данными источника.

5. Обсуждение результатов

Результаты проведенного исследования выходят за рамки простого технического сравнения моделей и ставят фундаментальные вопросы о природе данных и процессе интерпретации в цифровую эпоху. Главный вывод заключается в том, что получаемый с помощью больших языковых моделей структурированный датасет – это не объективное отражение источника, а продукт его целенаправленной интерпретации. В терминологии, предложенной Дж. Друкер, это не столько *data* (данные), сколько *capta* – данные, отобранные, «схваченные» через призму определенной аналитической оптики под конкретную исследовательскую задачу [3],[4],[19],[30]. Использование термина *capta* здесь не является простой заменой слова «данные». Он призван подчеркнуть важный нюанс, возникающий именно при работе с языковыми моделями. При традиционном исследовании отбор сведений (*capta*) осуществляется непосредственно историком, его критерии отбора являются частью профессиональной экспертизы. При использовании больших языковых моделей между исследователем и текстом появляется активный посредник – сама модель. Ее архитектура, а также заданный исследователем набор инструкций (промпт) и строгая схема классификации (онтология) и предопределяют, что именно будет «схвачено». Благодаря этому процесс извлечения данных становится технически воспроизводимым, а интерпретационная рамка исследователя из умозрительной концепции превращается во внешний, проверяемый и многократно используемый инструмент анализа. Применение этого единого инструмента к разным моделям со всей наглядностью демонстрирует их собственную интерпретационную природу. Обнаруженные фундаментальные различия в характере извлечения информации служат опровержением «иллюзии объективности» [4, с. 92-93] цифровых методов и доказывают, что автоматически полученные структурированные данные являются не пассивным отражением источника, а активным продуктом его машинной интерпретации. Каждая модель фактически реализует собственную аналитическую стратегию, исходя из своей архитектуры и корпуса данных, на которых она обучена.

На наш взгляд, этот вывод напрямую соотносится с концепцией машинной, или «цифровой», герменевтики, согласно которой автоматическое извлечение формализованной информации является не просто технически нейтральной операцией, а специфическим видом интерпретации [4],[26]. Безусловно, этот процесс имеет

неоспоримую техническую основу, однако он не сводится только к ней, поскольку результат зависит не только от настроек, заданных исследователем (промпт, онтология), но и от архитектурных особенностей и данных, на которых обучалась модель. Каждая из них предлагает свою собственную, внутренне последовательную версию структуризации текста. И ценность этого спектра прочтений заключается не в том, чтобы выбрать одно «правильное» и отвергнуть остальные, а в том, чтобы использовать их как аналитические инструменты, подсвечивающие разные грани источника. Так, исследователь, решающий задачу из области истории эмоций, может предпочесть модель, более чувствительную к эмоциональной составляющей повествования (как Grok3 в нашем случае), в то время как для реконструкции неявного исторического контекста более полезным окажется инструмент, нацеленный на извлечение косвенных свидетельств (Gemini-2.5-Pro). Этот подход не подменяет финальный синтез, который остается прерогативой историка, а вооружает исследователя новыми инструментами для анализа. Он позволяет осознанно выбрать модель под конкретную задачу и, рассмотрев источник с разных сторон, построить собственную, более полную и аргументированную интерпретацию.

На практическом уровне ключевым результатом стало выявление обратной зависимости между полнотой и точностью. Так, модель Gemini-2.5-Pro, обеспечивающая максимальный охват, является оптимальным инструментом для разведочного анализа и формирования гипотез, позволяя выявить ранее неочевидные связи. Однако цена такой полноты – высокий уровень «шума», требующий тщательной ручной верификации. Напротив, модели Deepseek-v3 и o3, ориентированные на высокую точность, лучше подходят для создания выверенных баз данных по заранее определенным параметрам, но с риском упустить важные нюансы и контекстуальные связи.

Из этого следует, что уже выбор конкретной языковой модели становится первым этапом конструирования *capta* – целенаправленным исследовательским актом, предопределяющим, какая именно информация будет извлечена из текста. Это наблюдение актуализирует в цифровой среде классическую дилемму исторического ремесла: противоречие между источником-ориентированным и проблемно-ориентированным подходами [4, с. 89],[36]. Первый, восходящий к принципу *ad fontes* («к источникам»), нацелен на максимально полный сбор всех свидетельств, рискуя утонуть в нерелевантных деталях. Эту стратегию воспроизводит Gemini-2.5-Pro с ее стопроцентной полнотой и высоким уровнем «шума». Второй подход, проблемно-ориентированный (*ceteris paribus*), предполагает сознательное ограничение оптики для ответа на конкретный вопрос, с риском пропустить данные, не вписывающиеся в первоначальную гипотезу. Эту логику отражает Deepseek-V3 с ее высокой точностью и значительным числом пропущенных косвенных свидетельств. Эта дилемма имеет и прямое следствие для интерпретации стандартных метрик оценки: если в технических задачах высокая точность часто является приоритетом, то для исторического анализа максимальная полнота в ряде случаев может быть предпочтительнее. Следовательно, языковые модели не устраняют эту фундаментальную дилемму, а лишь обостряют ее, требуя от историка осознанного выбора исследовательской стратегии.

Но несмотря на все эти методологические сложности, пожалуй, наиболее значимый практический результат для анализа эго-документов заключается в том, что современные модели, особенно Gemini-2.5-Pro, продемонстрировали способность к тонкой семантической декомпозиции. Разделение новостного факта, его эмоциональной оценки, рациональной рефлексии и идеологического предубеждения в рамках одного текстового фрагмента открывает новые горизонты для микроисторического анализа и истории эмоций. Это согласуется с выводами недавних исследований, демонстрирующих потенциал языковых моделей к выявлению субъективных пластов в дневниках.

В то же время исследование выявило и серьезные ограничения, особенно в способности

моделей работать со сложными косвенными свидетельствами. Наши выводы находят прямое подтверждение в результатах Т. Хильтманна, который в своем эксперименте с ChatGPT-4 и средневековой хроникой показал, что модель, успешно справляясь с общим анализом текста, допускает грубые фактические ошибки при попытке разобраться в сложных генеалогических связях баварских герцогов XIII века [26, S. 216-218]. В обоих случаях мы наблюдаем один и тот же феномен: модель уверенно оперирует частотными историческими фактами, но терпит неудачу там, где требуется гранулярное знание, редко встречающееся в обучающих данных.

Этот вывод подкрепляется и более широкими тестами: недавнее исследование показало, что исторические знания современных языковых моделей поверхностны и не достигают экспертного уровня [24]. В рамках теста HiST-LLM, основанного на базе данных по всемирной истории, точность лучших моделей едва достигла 46%, что свидетельствует о серьезных пробелах в знаниях. Примечательно, что модели парадоксальным образом показывают более высокую точность в отношении древних эпох, но хуже справляются со сложностью и объемом данных по более поздним историческим периодам. Кроме того, их знания оказались географически несбалансированными: модели лучше всего справляются с историей Северной и Латинской Америки и хуже всего – с историей Океании и Африки южнее Сахары. Наши результаты подтверждают этот вывод и конкретизируют его, показывая специфические паттерны ошибок при работе с эго-документами. Проблема «галлюцинаций», особенно заметная в поле `brief_context`, свидетельствует, что без обязательной ручной верификации со стороны профессионального историка автоматизация неизбежно будет приводить к искажениям. В этом контексте роль историка, использующего инструмен расширяется: к традиционным навыкам поиска и критики источников добавляется новая компетенция – критическая оценка и интерпретация результатов машинного анализа.

Следует учитывать, что представленная работа обладает определёнными ограничениями. Во-первых, она основана на анализе единственного эго-документа. Выводы исследования не могут быть автоматически перенесены на другие жанры, эпохи и языки. Производительность языковых моделей существенно варьируется в зависимости от языка, исторического периода и типа документов [25]. Во-вторых, данное исследование представляет собой лишь срез на определенный момент времени в стремительно развивающемся мире больших языковых моделей. В-третьих, создание «золотого стандарта» для оценки само по себе является актом экспертной интерпретации, что вносит элемент субъективности в количественные метрики.

6. Заключение

Настоящее исследование на материале дневника К.А. Березкина позволило количественно и качественно сравнить возможности современных больших языковых моделей в задачах автоматического извлечения структурированной информации. Наш опыт показал, что большие языковые модели способны не только извлекать события и имена, но и распутывать сложные семантические слои текста: отделять источник информации от личного мнения, фиксировать эмоциональные оценки и улавливать моменты саморефлексии автора. Этот аналитический потенциал открывает перспективные возможности для реконструкции микроисторического контекста на качественно новом уровне детализации. Данный подход позволяет значительно оптимизировать работу с объемными источниками личного происхождения, автоматизируя трудоемкий процесс первоначального отбора релевантной информации из большого массива текста.

Результаты сравнительного анализа однозначно показывают, что ни одна из протестированных моделей не является универсальным инструментом. Выявленная

обратная зависимость между полнотой и точностью ставит историка перед необходимостью осознанного методологического выбора: либо исчерпывающий, но «шумный» охват для разведочного анализа, либо высокоточное, но потенциально неполное извлечение для создания выверенных баз данных. Таким образом, взаимодействие с языковыми моделями становится не просто процессом механического извлечения «объективных» фактов. Искусственный интеллект выступает как ассистент, способный предложить множественные структурированные прочтения источника. Историк же, используя эти прочтения как эвристический инструмент – своего рода «цифровую лупу», – не просто проверяет факты, но получает возможность для более глубокого анализа, сохраняя за собой неотъемлемую прерогативу финальной интерпретации и синтеза.

В целом проведённая работа подтверждает: интеграция искусственного интеллекта в историческое исследование требует симбиоза машинных мощностей с экспертной оценкой профессионального историка. Дальнейший прогресс в этой области связан не столько с созданием все более мощных моделей, сколько с развитием у историков навыков «критической цифровой грамотности» – умения не просто использовать инструмент, а понимать его логику, предвидеть ошибки и осмысленно интегрировать его результаты в свое исследование. Это означает переход от роли пользователя цифровых инструментов к роли архитектора цифрового исследования, способного осознанно формировать и верифицировать данные, полученные в диалоге с искусственным интеллектом.

7. Доступность данных и кода

Все данные, исходный код, онтология и результаты, лежащие в основе данного исследования, находятся в открытом доступе в репозитории GitHub. Репозиторий включает:

- Python-скрипт для извлечения информации (src/);
- исходный текст дневника и эталонный набор данных (data/);
- онтологию в формате JSON (knowledge_map.json);
- полные выходные файлы протестированных моделей (results/model_outputs/);
- приложения к статье (docs/).

Все материалы доступны по адресу:
https://github.com/alexeyvkuznetsov/diary_event_extraction/

Библиография

1. Авербух Р. А. Царская интервенция в борьбе с венгерской революцией 1848–1849. – Москва : Соцэкгиз, 1935. – 232 с.
2. Березкин К. А. Дневник [Электронный ресурс] / подгот. текста: А. Завина, А. Микерин // Электронный корпус "Прожито". – URL: <https://corpus.prozhito.org/person/925> (дата обращения: 08.09.2025).
3. Володин А. Ю. Исторические исследования в контексте датаизма: методологический аспект // Вестник Пермского университета. История. – 2023. – № 4(63). – С. 135-147. – DOI: 10.17072/2219-3111-2023-4-135-147. – EDN: GFSIOJ.
4. Володин А. Ю. Цифровая герменевтика исторического источника: формализация как толкование // Вестник Пермского университета. История. – 2025. – № 2(69). – С. 87-100. – DOI: 10.17072/2219-3111-2025-2-87-100. – EDN: PWRHKI.

5. Гросул В. Я. Русские участники зарубежных революций первой половины XIX века // Новая и новейшая история. – 2007. – № 4. – С. 21-40. – EDN: IAMEKF.
6. Кузнецов А. В. Большие языковые модели как инструмент историка // Новые информационные технологии в образовании и науке. – 2024. – № 13. – С. 75-92. – DOI: 10.17853/2587-6910-2024-13-75-92. – EDN: QDSUHV.
7. Кузнецов А.В. За пределами тематического моделирования: анализ исторического текста с помощью больших языковых моделей // Историческая информатика. 2024. № 4. С. 47-65. DOI: 10.7256/2585-7797.2024.4.72560 EDN: UOIKPJ URL: https://nbpublish.com/library_read_article.php?id=72560
8. Линькова Е. В. Внешнеполитический курс Николая I в оценках отечественных консерваторов середины XIX в. // Вестник РУДН. Серия: История России. – 2014. – № 1. – С. 17-23. – EDN: RZVVTB.
9. Медведева О. Э. Николай I и европейские революции в историографии // Вестник научной ассоциации студентов и аспирантов исторического факультета Пермского государственного гуманитарно-педагогического университета. Серия: Studis historica juvenum. – 2017. – № 1 (13). – С. 132-136.
10. Международные отношения на Балканах, 1830–1856 гг. / [Г. Л. Арш, И. С. Варта, В. Н. Виноградов и др.] ; отв. ред. В. Н. Виноградов ; АН СССР, Ин-т славяноведения и балканистики. – Москва : Наука, 1990. – 349 с.
11. Мещерякова А. О. Русские консерваторы и европейские революции 1848–1849 годов // Тетради по консерватизму. – 2017. – № 2. – С. 61-67. – EDN: YMFTLR.
12. Орлик И. И. Венгерская революция 1848–1849 годов и Россия // Новая и новейшая история. – 2008. – № 2. – С. 21-40. – EDN: IRGMBF.
13. Орлик О. В. Передовая Россия и революционная Франция (первая половина XIX века). – Москва : Наука, 1973. – 299 с.
14. Стыкалин А. С. Правда и вымысел о реакции российского общества на венгерскую революцию 1848 г. Венгерская кампания 1849 г. и капитан Гусев // Историческая экспертиза. – 2014. – № 1. – С. 38-54. – EDN: WCHJSL.
15. Федоров А. В. Отношение передовых людей России к Венгерской революции 1848–1849 годов // Вопросы истории. – 1957. – № 2. – С. 89-96.
16. Alhamed F., Ive J., Specia L. Using large language models (LLMs) to extract evidence from pre-annotated social media data // Proceedings of the 9th Workshop on Computational Linguistics and Clinical Psychology (CLPsych 2024). – 2024. – P. 232-237.
17. Bernard G. Détection et suivi d'événements dans des documents de presse historiques : дис. – Université de La Rochelle, 2022.
18. De Toni F. et al. Entities, Dates, and Languages: Zero-Shot on Historical Texts with T0 // BigScience 2022-International Workshop on Challenges & Perspectives in Creating Large Language Models 2022. – 2022. – P. 75-83.
19. Drucker J. Humanities Approaches to Graphical Display [Электронный ресурс] // Digital Humanities Quarterly. – 2011. – Vol. 5, № 1. – URL: <http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html> (дата обращения: 08.09.2025).
20. Gao T., Langlais P. RaTE: a Reproducible automatic Taxonomy Evaluation by Filling the Gap // Proceedings of the 15th International Conference on Computational Semantics (IWCS 2023). – 2023. – P. 173-182.
21. Goloviznina V. et al. Do LLMs Understand Why We Write Diaries? A Method for Purpose Extraction and Clustering // arXiv preprint arXiv:2506.00985. – 2025. – URL: <https://arxiv.org/abs/2506.00985> (дата обращения: 08.09.2025).
22. González-Gallardo C. E. et al. Leveraging open large language models for historical named entity recognition // International Conference on Theory and Practice of Digital

- Libraries. – Cham : Springer Nature Switzerland, 2024. – P. 379-395.
23. Han R. et al. An empirical study on information extraction using large language models // arXiv preprint. – 2023. – arXiv:2305.14450. – URL: <https://arxiv.org/abs/2305.14450> (дата обращения: 08.09.2025).
24. Hauser J. et al. Large Language Models' Expert-level Global History Knowledge Benchmark (HiST-LLM) // Advances in Neural Information Processing Systems. – 2024. – Т. 37. – P. 32336-32369.
25. Hiltmann T. et al. NER4all or Context is All You Need: Using LLMs for low-effort, high-performance NER on historical texts. A humanities informed approach // arXiv preprint. – 2025. – URL: <https://arxiv.org/abs/2502.04351> (дата обращения: 08.09.2025).
26. Hiltmann T. Hermeneutik in Zeiten der KI: Large Language Models als hermeneutische Instrumente in den Geschichtswissenschaften // KI:Text: Diskurse über KI-Textgeneratoren / ed. by G. Schreiber, L. Ohly. – Berlin ; Boston : De Gruyter, 2024. – S. 201-232.
27. König M. ChatGPT und Co. in den Geschichtswissenschaften – Grundlagen, Prompts und Praxisbeispiele [Электронный ресурс] // Digital Humanities am DHIP. – 2024. – 19 Aug. – URL: <https://dhdhi.hypotheses.org/9197> (дата обращения: 08.09.2025).
28. Laato J. et al. Extracting Social Connections from Finnish Karelian Refugee Interviews Using LLMs // arXiv preprint. – 2025. – arXiv:2502.13566. – URL: <https://arxiv.org/abs/2502.13566> (дата обращения: 08.09.2025).
29. Lai V. D. et al. Event extraction from historical texts: A new dataset for black rebellions // Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. – 2021. – P. 2390-2400.
30. Lavin M. Why digital humanists should emphasize situated data over capta [Электронный ресурс] // Digital Humanities Quarterly. – 2021. – Vol. 15, № 2. – URL: <http://www.digitalhumanities.org/dhq/vol/15/2/000556/000556.html> (дата обращения: 08.09.2025).
31. Li J. et al. DiaryHelper: Exploring the use of an automatic contextual information recording agent for elicitation diary study // Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. – 2024. – P. 1-16.
32. Pan H. et al. Taxonomy-Driven Knowledge Graph Construction for Domain-Specific Scientific Applications // Findings of the Association for Computational Linguistics: ACL 2025. – 2025. – C. 4295-4320.
33. Piskorski J., Yangarber R. Information extraction: Past, present and future // Multi-source, multilingual information extraction and summarization. – Berlin, Heidelberg : Springer Berlin Heidelberg, 2012. – P. 23-49.
34. Santini C. Combining language models for knowledge extraction from Italian TEI editions // Frontiers in Computer Science. – 2024. – Т. 6. – P. 1472512.
35. Shin D. et al. Using large language models to detect depression from user-generated diary text data as a novel approach in digital mental health screening: instrument validation study // Journal of Medical Internet Research. – 2024. – Т. 26. – P. e54617. – DOI: 10.2196/54617. – EDN: VFSRAP.
36. Thaller M. Historical Information Science: Is There such a Thing? New Comments on an old Idea // Seminario Discipline Umanistiche e Informatica. Il Problema dell'Integrazione/Ed. T. Orlandi. Roma. – 1993. – P. 51-86.
37. Tang X. et al. CHisIEC: An Information Extraction Corpus for Ancient Chinese History // Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). – 2024. – P. 3192-3202.
38. Xie Y., La Mela M., Tell F. Multimodal LLM-assisted Information Extraction from Historical Documents: The Case of Swedish Patent Cards (1945–1975) and ChatGPT // The 9th Digital Humanities in the Nordic and Baltic Countries Conference (DHNB 2025), March 5-

7, 2025, Tartu, Estonia. – University of Oslo Library, 2025. – P. 1-15.

39. Xu D. et al. Large language models for generative information extraction: A survey // *Frontiers of Computer Science*. – 2024. – Т. 18. – № 6. – P. 186357.

40. Zhang Y. et al. Teleclass: Taxonomy enrichment and llm-enhanced hierarchical text classification with minimal supervision // *Proceedings of the ACM on Web Conference 2025*. – 2025. – P. 2032-2042.

Результаты процедуры рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.

Со списком рецензентов издательства можно ознакомиться [здесь](#).

Рецензируемая статья относится ко все более увеличивающемуся потоку материалов, посвященных тем или иным аспектам использования искусственного интеллекта (ИИ) в исследовательской работе. Автор концентрируется на извлечении исторической информации из текстовых источников, рассматривая источник личного происхождения. Предметом изучения является разработка методологии автоматического извлечения структурированной информации из дневниковых записей с помощью больших языковых моделей и формирование набора данных для решения конкретных исторических задач. В данном случае речь идет о восприятии европейских революций 1848–1849 гг. в российской провинции на основе анализа дневниковых записей вологодского гимназиста К.А. Березкина.

Актуальность исследований, которым посвящена статья, не требует подробной мотивации вследствие огромного интереса и внимания широкой научной общественности буквально ко всем направлениям, связанным с применением технологий ИИ. Что касается научной новизны, то она обусловлена, с одной стороны, продвижением в решении сложной «глубинной» задачи исторического источниковедения по извлечению информации из сложного по своей структуре и субъективности источника, с другой – формированием новой модели для подобных исследований.

Статья структурирована по современным канонам создания научных текстов. Она начинается с введения, где формулируется проблематика, цели и задачи исследования. Во второй части статьи проведен обзор литературы, в котором на основе изучения и осмысления новейших отечественных и, преимущественно, зарубежных публикаций автором отмечаются некоторые пробелы в существующих исследованиях, которые восполняются в настоящей статье. В следующем, третьем разделе, дается краткая характеристика источника и обосновывается методика работы с ним, включая разработку онтологии и инструментов исследования, выбор больших языковых моделей и процедуру извлечения информации. В четвертом разделе приводятся результаты анализа, их оценка, апробация и обсуждение результатов. Далее подводятся итоги исследования и дается ссылка на все доступные материалы, положенные в основу статьи.

Библиография статьи насчитывает 39 позиций, где помимо работ, связанных с историческим контекстом исследования, значительное место занимают новейшие материалы, связанные с привлечением в исторические исследования методов и технологий ИИ. Статья содержит 3 таблицы, иллюстрирующих разные ступени проведенного исследования. Хотелось бы сделать замечание по поводу их оформления: заголовки таблиц принято размещать над таблицами, а не под ними. Также в тексте отсутствует прямая отсылка на таблицу 3.

Рецензируемая статья ставит сложные вопросы методологии использования искусственных нейросетей при изучении исторических текстов и в то же время

затрагивает проблемы историографического анализа. В этой связи автор отмечает, что работа решает «комплексную задачу по разработке и тестированию подхода, позволяющего трансформировать неструктурированный текст источника в датасет, пригодный для решения конкретной историографической задачи – анализа восприятия европейских революций 1848–1849 гг. в российской провинции». Материалом для рассмотрения обеих задач послужил дневник вологодского гимназиста К.А. Березкина за 1849 год, оцифрованный текст которого размещен в цифровом ресурсе «Прожито». Автор отмечает, что «работа с подобным источником сопряжена со значительными методологическими сложностями: прямые упоминания о революционных событиях редки и вплетены в повседневную жизнь» гимназиста, а центральное место в дневнике «занимают учебные будни в гимназии, светская жизнь провинциального дворянства, а также глубокая личная рефлексия автора о собственных поступках, чувствах и планах на будущее». Казалось бы, в такой ситуации исследователю предстоит непростая задача – вникая в дневниковые записи, владея контекстом революций 1848-1849 гг., разобраться с переплетениями рефлексий о повседневной жизни и откликами на революционные события. Чем здесь может помочь искусственный интеллект? Автор статьи должен дать более четкий ответ на этот вопрос, тем более что из 249 дневниковых записей всего 34 (выявленных автором статьи «вручную») имеют прямое или косвенное отношение к революционной тематике.

Отметим, что сформулированная автором задача «анализа восприятия европейских революций 1848–1849 гг. в российской провинции» не может основываться на дневниковых записях вологодского гимназиста. Решение такой задачи потребовало бы исследования других, более репрезентативных источников по различным губерниям. Поэтому представляется, что историографическую задачу следует снять из поставленных автором задач данной статьи. По сути, в статье предложена методика сопоставления возможностей широко известных больших языковых моделей (LLM): Gemini-2.5-Pro, o3 – модели от OpenAI, Grok-3, Deepseek-V3, «извлечения структурированной информации из дневника К.А. Березкина» и оценки их эффективности «для формирования качественного набора данных (датасета)». Эта задача является значимой сама по себе. Дневник в данном случае надо рассматривать в качестве «полигона» (термин автора статьи) для апробации возможностей LLM. Следует отметить, что оценка производительности 4-х рассмотренных моделей производится на основе трех метрик, используемых в информационных системах и позволяющих получить количественные критерии при сопоставлении этих моделей.

В первом разделе статьи автор отмечает, что он стремится «внести вклад в методологическую дискуссию о статусе автоматически извлеченных данных и роли историка в эпоху “цифровой герменевтики”». Большинство историков не знают, что наступила «эпоха цифровой герменевтики». Возможно, этот термин пока рано продвигать в профессиональной исторической среде.

В статье отмечается, что ключевой вывод исследования состоит в том, «что автоматизированное извлечение является не технической операцией, а формой цифровой герменевтики». Думается, что автоматизированные процедуры все-таки не могут не быть в значительной мере техническими. Требуется пояснения следующий тезис автора: «итоговый датасет – это не объективные данные (data), пассивно «обнаруженные» в источнике, а *carpa* – набор сведений, отобранных под конкретную задачу». Но ведь, обращаясь к тексту источника, мы задаемся тем или иным содержательным вопросом для решения той или иной конкретной задачи; например, об отражении отношения автора к революциям, или к его друзьям по гимназии. Именно так историк и работает с источником. Что добавляет здесь использование слова *carpa*?

Еще один вопрос возникает в связи с выявленной фактической ошибкой в

сгенерированным одной из LLM рассуждении. Автор справедливо отмечает, что этот пример «служит важнейшим методологическим предостережением: даже самая продвинутая модель может генерировать правдоподобные, но фактически неточные контекстуальные справки», что «повышает требования к критической экспертизе и эрудиции самого историка, который остается финальным арбитром в интерпретации источника и его контекста». Но ведь разные LLM, как показано в статье, генерируют описания упомянутых в тексте источника событий по-разному, с разной степенью аналитичности, вплоть до «упрощенного перефразирования исходного текста». Следовательно, при использовании нескольких LLM, историк столкнется с различными машинными интерпретациями. Каковы должны быть действия историка в этой ситуации? По мнению автора статьи, «каждая модель предлагает свою собственную, внутренне последовательную версию структуризации текста. В результате исследователь потенциально может получить не единственно верный «объективный» результат, а целый спектр возможных прочтений источника, что лишь подчеркивает его собственную роль как финального арбитра». Думается, в этой ситуации историк предпочтет дать свою интерпретацию (если он имеет должную квалификацию, а не полагается в основном на искусственный интеллект).

Говоря об источнико-ориентированном и проблемно-ориентированном подходах, следует дать ссылку на автора этой терминологии – проф. Манфреда Таллера (Кёльнский университет).

Требует уточнения и еще один тезис автора статьи, считающего, что «применение искусственного интеллекта повышает требования к критической экспертизе историка, смещая его роль от поиска информации к верификации и интерпретации машинных результатов». Вообще-то роль историка в поиске релевантных источников вряд ли заменит LLM, но вот видеть себя в роли верификатора и интерпретатора машинных результатов согласятся немногие из профессиональных историков. В этой связи можно процитировать автора статьи, отмечающего, что недавнее исследование показало, что исторические знания современных языковых моделей поверхностны и не достигают экспертного уровня. Однако это положение, видимо, временное: генеративные нейросети развиваются, их возможности растут.

Можно только согласиться с выводами автора об ограничениях представленной работы: «она основана на анализе единственного эго-документа», а полученные результаты не могут быть автоматически перенесены на другие жанры и эпохи. Представляется, что статья выиграет, если будет содержать пример сформированного датасета (его фрагмент).

В целом статья вносит существенный вклад в проблематику применения технологий ИИ в исторических исследованиях, изложенная в тексте методика сравнения эффективности рассмотренных генеративных нейросетей может быть применена в других исследованиях исторических текстов. При учете сделанных замечаний и пожеланий рецензируемая статья, соответствующая профилю и стилю журнала «Историческая информатика», вызовет интерес читательской аудитории. Статья рекомендуется к доработке, после которой она может быть опубликована.

Результаты процедуры повторного рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.

Со списком рецензентов издательства можно ознакомиться [здесь](#).

Предметом рецензируемого исследования является разработка и апробация методологии автоматического извлечения структурированной информации из сложного

эго-документа (личного дневника) с использованием больших языковых моделей (БЯМ). В качестве конкретного кейса авторы избрали задачу выявления восприятия европейских революций 1848–1849 гг. в российской провинции на материале дневника вологодского гимназиста. Это соответствует заявленной цели – не просто техническому сравнению моделей, а решению конкретной исследовательской проблемы с помощью цифровых методов.

Методологическая основа исследования представляется проработанной и комплексной. Авторы не ограничиваются простым прогоном текста через API моделей, а выстраивают целостный исследовательский контур: от оцифровки источника и разработки доменно-специфической онтологии, задающей структуру для извлечения данных, до тщательного инжиниринга промптов и детальной схемы валидации выходных данных. Ключевым методологическим решением является использование онтологии как формализованной модели предметной области, что позволяет минимизировать субъективность и снизить вероятность галлюцинаций моделей. Выбор четырех флагманских моделей (Gemini-2.5-Pro, o3, Grok3, Deepseek-V3) для сравнительного анализа оправдан и отражает стремление авторов охватить все ведущие архитектуры. Методика оценки на основе «золотого стандарта» с расчетом стандартных метрик (Precision, Recall, F1-score) дополнена качественным анализом, что позволяет дать многомерную оценку эффективности каждой модели. Такой подход демонстрирует высокую исследовательскую культуру и соответствие современным стандартам цифровых гуманитарных исследований.

Актуальность работы не вызывает сомнений. Использование БЯМ для анализа исторических текстов, особенно личных нарративов, является одним из перспективных и дискуссионных направлений сегодня. Статья отвечает на настоятельную потребность в методических разработках, которые бы не просто демонстрировали возможности технологий, но и предлагали критический, выверенный инструментарий для их применения в конкретных исторических исследованиях. Обращение к проблеме восприятия глобальных событий на микроуровне провинциальной жизни также вписывается в тренд на микроисторическую оптику.

Научная новизна статьи выражена отчетливо. Авторы справедливо указывают на два пробела в литературе: отсутствие исследований по комплексному извлечению разных типов информации из исторических дневников и отсутствие систематического сравнительного анализа производительности ведущих БЯМ на таком материале. Исследование вносит вклад в заполнение этих двух лакун. Новизна заключается также в разработанной авторами методологии, сочетающей онтологический подход с продвинутым инжинирингом промптов, и в полученных результатах, которые выявляют не только количественные, но и качественные, сущностные различия в аналитических стратегиях моделей (например, в способности к семантической декомпозиции текста).

Статья написана ясным, научным стилем, с соблюдением всех формальных требований. Структура логична и традиционна для научной работы. Изложение последовательное и аргументированное. Текст хорошо сбалансирован между техническими деталями (которые необходимы для воспроизводимости) и исторической интерпретацией. Визуализация в виде таблиц и примеров JSON-объектов помогает восприятию сложного материала. Единственное пожелание – немного сократить раздел с детальным описанием полей JSON-схемы, перенеся часть этих деталей в приложение, чтобы не перегружать основной текст.

Авторы продемонстрировали глубокое знакомство с современной литературой как по цифровым гуманитарным наукам и компьютерной лингвистике (включая свежие препринты 2024-2025 гг.), так и по историографии революций 1848-1849 гг. и николаевской России. Библиография включает ключевые работы по информационному

извлечению, онтологиям, применению ИИ в истории, а также фундаментальные исторические исследования. Список составлен корректно, с соблюдением всех библиографических стандартов.

Авторы предвосхищают возможную критику, что является признаком зрелой работы. Они открыто обсуждают ограничения своего исследования: работа с одним документом, быстрое устаревание результатов в связи с развитием моделей, элемент субъективности при создании «золотого стандарта». Главный контраргумент потенциальным оппонентам, которые могут скептически относиться к автоматизированному анализу, заключается в том, что авторы не предлагают заменять историка, а предлагают новый мощный инструмент для его работы, требующий не меньшей, а возможно и большей критической экспертизы. Тезис о том, что результат извлечения – это не объективные «данные» (data), а «исследовательский улов» (capta), является сильной методологической позицией, снимающей множество потенциальных вопросов.

Исследование представляет собой не просто эксперимент, а целостный, воспроизводимый методологический конвейер – от сырого текста до интерактивного дашборда. Авторы выходят за рамки простого сравнения метрик, проводя тонкий качественный анализ «слепых зон» и аналитических предпочтений каждой модели, что крайне ценно для понимания их природы. Публикация кода, данных, онтологии и результатов позволяет другим исследователям не только проверить выводы исследования, но и использовать предложенные наработки в своих проектах, что значительно повышает ценность работы для научного сообщества.

Как и отмечают сами авторы, выводы, сделанные на материале одного дневника, хоть и репрезентативного, нуждаются в проверке на других текстах иных жанров, эпох и авторства, что, конечно, сужает область обобщения сделанных выводов. Несмотря на все достоинства метода, эталонная разметка неизбежно несет на себе отпечаток экспертного восприятия самих исследователей. В идеале для валидации подобных методов хорошо бы иметь несколько независимых экспертных разметок для расчета коэффициента согласия (например, Cohen's kappa). Конкретные количественные результаты по эффективности моделей могут быстро измениться с выходом их новых версий (некоторые новые версии уже появились). Однако методологическая рамка и качественные выводы о различии стратегий извлечения сохраняют свою актуальность.

Ключевой вывод статьи о том, что БЯМ выступают не нейтральным инструментом извлечения, а активным интерпретатором, генерирующим «capta», а не «data», является фундаментальным и важным. Это меняет сам подход к работе с этими технологиями: от слепого доверия к критическому диалогу. Авторы убедительно показывают, что выбор модели – это уже методологический и даже герменевтический акт, определяющий стратегию исследования. Этот вывод имеет далеко идущие последствия для цифровых гуманитарных исследований. Возможно, в дальнейшем имеет смысл сопоставить предложенный подход с RAG-подходом (retrieval-augmented generation или генерацией, дополненной выборкой), выяснив какую долю успеха составляет онтологический подход, или им возможно пренебречь, а достаточно нормализовать исходные данные.

Статья вызовет значительный интерес у широкого круга специалистов. В первую очередь, у историков, занимающихся цифровыми методами, специалистов по исторической информатике и digital humanities. Она также будет полезна лингвистам и компьютерным специалистам, работающим в области обработки естественного языка и информационного извлечения, особенно на основе исторических материалов. Наконец, статья может заинтересовать историков, изучающих имперский период России, как пример нового подхода к анализу знакомых источников.

Настоящая статья соответствует проблематике журнала «Историческая информатика», отличается научной новизной, методологической строгостью, практической значимостью,

богатой библиографией и корректным оформлением. Статья вносит вклад в развитие методологии применения искусственного интеллекта в исторических исследованиях и задает важный тон для дискуссии в будущих работах в этой области. Незначительные замечания, высказанные в рецензии, носят рекомендательный характер и могут быть учтены авторами по их усмотрению.

На основании всего вышеизложенного, считаю, что статья «Автоматическое извлечение информации из эго-документа: сравнительный анализ эффективности больших языковых моделей на примере дневника К.А. Березкина» заслуживает публикации в журнале «Историческая информатика».