

## DIGITAL HUMANITIES: СКВОЗЬ АЛГОРИТМЫ К ЗНАНИЯМ

УДК 930.23

doi 10.17072/2219-3111-2025-2-87-100

EDN: PWRHKI

ASJC 1202

ГРНТИ 03.01.07

Ссылка для цитирования: 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



### ЦИФРОВАЯ ГЕРМЕНЕВТИКА ИСТОРИЧЕСКОГО ИСТОЧНИКА: ФОРМАЛИЗАЦИЯ КАК ТОЛКОВАНИЕ

**А. Ю. Володин**

Московский государственный университет имени М. В. Ломоносова, 119192, Россия, Москва, Ломоносовский пр-т, 27, корп. 4, Г-423;

Институт цифровых гуманитарных исследований Сибирского федерального университета, 660130, Россия, Красноярск, Свободный пр-т, 82, стр. 1, ауд. 4-60  
voldin@hist.msu.ru

Researcher ID: J-9603-2012

Scopus Author ID: 5721905252

SPIN-код: 9627-6687

Статья посвящена осмыслиению трансформации методологии исторической науки под влиянием «цифрового поворота», с фокусом на ключевую дисциплину – источниковедение. Автор исследует, как стремительное развитие технологий искусственного интеллекта (ИИ), особенно машинного обучения, ставит перед исследователями-историками новые вызовы и открывает возможности, требуя переосмыслиения традиционных практик работы с историческими источниками. Центральной проблемой статьи является процесс перехода от исторического источника к формализованным данным (например, путем цифровой герменевтики) и обратно к интерпретации и новому историческому знанию. Автор подчеркивает, что этот процесс не просто техническая операция, а глубоко методологический акт абстрагирования (на примере собранных исторических данных, *carta*), сопряженный с неизбежными потерями нюансов и требующий критической рефлексии. Рассматриваются методологические дилеммы исторического исследования, такие как противостояние источнико-ориентированного (*ad fontes*, стремление к максимальной полноте) и проблемно-ориентированного (*ceteris paribus*, фокус на выбранных параметрах) подходов, их эволюция в цифровую эпоху и связанные с ними дискуссии. Критически анализируется «иллюзия объективности» цифровых методов. Автор утверждает, что алгоритмы, цифровые инструменты (СУБД, ГИС, NLP) и интерфейсы не нейтральны: они несут в себе предубеждения, ограничивают фокус, упрощают сложность и иногда даже предопределяют исследовательские вопросы, что порождает риски: предвзятость алгоритмов, селективность данных, влияние дизайна визуализаций и интерфейсов, некритичное принятие результатов ИИ. В качестве перспективного решения, преодолевающего ключевые ограничения генеративных моделей (галлюцинации, «черный ящик», оторванность от источников), в статье подробно рассматривается RAG-подход (Retrieval-Augmented Generation), показаны его преимущества для историка. Главный вывод статьи подчеркивает незаменимую роль историка-исследователя в эпоху ИИ. RAG-системы и подобные инструменты выступают не как замена, а как мощные «цифровые лупы» или интеллектуальные ассистенты, берут на себя трудоемкие задачи поиска и первичной обработки данных, освобождая время историка для критической интерпретации, оценки достоверности источников, построения аргументации, формулировки исследовательских вопросов и создания нового исторического знания.

**Ключевые слова:** данные, капта, герменевтика, формализация, историческая информатика, методология исторического исследования, цифровые гуманитарные исследования, цифровая история, RAG.

ἀνθρωπον, Φησί, Ζητῶ.  
Διογένης ο Κυνικός  
[Διογένης Λαέρτιος VI:41]

Ищу человека.  
Диоген Синопский  
[Диоген Лаэртский VI:41]

Мир науки никогда не будет прежним. Кажется, что мы говорим так слишком часто. Но успехи машинного обучения середины 2020-х гг. изумляют, пугают и вдохновляют одновременно. Как найти место для ученого-исследователя, вокруг которого ошеломительно быстро развиваются технологии искусственного интеллекта, рассмотрим на примере классического и, уверен, принципиально важного для любого историка предмета – источниковедения. Источниковедение неслучайно оказывается в центре внимания дискуссий о «цифровом повороте» в исторических исследованиях, потому что именно на основе источниковедения строится современная методология превращения исторических наблюдений в исторические данные. Исторические данные – представление сведений с исторической датировкой (или собранных за определенный период) в формализованном виде, пригодном для передачи, интерпретации или обработки людьми или компьютерами – оказывается основой для задач и решений машинного обучения. Теоретическое источниковедение особое внимание уделяет носителю исторической информации, источник – посредник, *medium*, средство коммуникации исторической информации – позволяет делать ученому опосредованные систематические наблюдения о прошлом, являющиеся основой научного исторического исследования.

Источниковедение сегодня главная методологическая основа исторического знания. Именно понимание источника как главной опоры для сбора исторических данных и наблюдений отличает историка в современном хоре цифровых гуманитарных наук. Как справедливо отметил С. О. Шмидт, «источниковедение возникло из практики» [Шмидт, 1997, с. 23]. Причем источниковед смотрит на источник в контексте современных источнику систем коммуникации, пытается расшифровать и оценить источник как с точки зрения его современников, так и «прочитать» его «свежими, и нынешними очами» [Там же, с. 46]. Следуя за правилами классических методик, практика исторического исследования в цифровую эпоху существенно меняется в инструментальном смысле. Историки постепенно овладевают навыками более проникновенного «прочтения» исторического источника, будь то мультиспектральная съемка или выявление скрытых связей путем сложной семантической разметки полнотекстовой базы данных. Соответственно, возникает то, что можно назвать более адекватным отражение содержания источников в познавательных образах [Там же]. Отметим, что сейчас не рассматриваем проблемы источниковедения цифровой эпохи. Источниковедение электронных исторических источников подробнейшим образом рассмотрено Ю. Ю. Юмашевой в пионерской монографии «Источниковедение информационной эпохи» [Юмашева, 2024]. Самостоятельным сюжетом современных исследований является анализ взаимосвязей исторического знания и философской герменевтики [Fairfield, 2025; Romele, 2020].

Одним из широко распространенных способов систематизации исторических наблюдений является их формализация с последующей фиксацией таких наблюдений в наборе данных, датасете. Фиксация наблюдений совсем не такой формальный процесс, она основывается на прочтении и толковании исследователя, зависит в том числе и от того, с каким вопросом историк приходит к источнику. Очевидно, что историк должен иметь представление о природе и особенностях исследуемых источников. Но в этом смысле природа источника не играет решающей роли, хотя с технологической точки зрения мы часто предполагаем, что цифровые копии исторических источников или источники, существующие в формате компьютерных файлов, должны легче поддаваться формализации и машинной обработке. На практике это оптимистическое ожидание совсем не всегда может оправдаться. И здесь уместно говорить про цифровую герменевтику как подход, позволяющий перевести наблюдение историка в машиночитаемый набор данных.

Цифровая герменевтика – перевод наблюдений историка, полученных из исторических источников, в данные. Можно назвать эту процедуру абстрагированием; при этом с методической точки зрения это формализация. Данные – это представление информации в формализованном виде, пригодном для передачи, связи или обработки людьми и машинами. Формализация может осуществляться разными способами: наблюдение может стать записью в базе данных, точкой с определенной координатой на карте, деталью объекта на трехмерной модели и т.д. Как отмечал К. Шёх, гуманитарные данные – это цифровая, выборочно сконструированная, управляемая компьютером абстракция, представляющая некоторые аспекты определенного объекта гуманитарного исследования. «Независимо от того, являемся ли мы историками, использующими тексты или другие культурные артефакты как окна в другое время или другую культуру, или же мы литературоведы, использующие знания других времен и культур для построения смысла текстов, – утверждает К. Шёх, – цифровые данные добавляют еще один уровень посредничества в уравнение. Данные (а также инструменты, с помощью которых мы ими манипулируем) добавляют сложности в отношения между исследователями и их объектами изучения» [Schöch, 2013, р. 3–4].

Методически данные получаются путем абстрагирования, т.е. отвлечения от несущественных (для исследователя и текущего исследования) характеристик объекта с целью определить его основные свойства и признаки, что позволяет сформулировать абстрактные понятия и модели, которые должны помочь понять суть изучаемого явления. При этом наблюдения собираются в определенную коллекцию (датасет), собранные сведения можно назвать каптой (capta), или исследовательским «уловом» из изученных исторических источников, полнота и представительность которого требуют дополнительного доказательства (подробнее см. [Володин, 2019; Drucker, 2011; Lavin, 2021]).

Абстрагирование предполагает отбрасывание конкретных особенностей опосредованно (т.е. через информацию исторического источника) наблюдаемого явления прошлого, позволяющее сосредоточиться на его основных и общих характеристиках. Исследователь оказывается в ситуации выбора: держаться за полноту и конкретность наблюдения (историки часто выступают против любой редукции, которая рассматривается как упрощение) или же конструировать модель явления прошлого, осознанно фокусируя внимание на определенных чертах изучаемого явления. Принцип первого подхода – *ad fontes* (к источникам) – предполагает сбор максимально большого количества сведений из источника (при этом, как нас учит теоретическое источниковедение, всех сведений мы не раздобыдем, а источник останется неисчерпаемым, по И. Д. Ковальченко [Ковальченко, 2003, с. 132]), именно так формулировалась цель источнико-ориентированного подхода, которую критиковал П. Доорн. Принцип второго подхода – *ceteris paribus* (при прочих равных) – основан на допущении, что исследователь должен отобрать интересные и важные черты и параметры явления прошлого, а все остальные параметры остаются постоянными или неизменными, но фактически находящимися за пределами исследовательского наблюдения. Принципы исторического моделирования представлены в фундаментальном исследовании «Моделирование исторических процессов: от реконструкции реальности к анализу альтернатив» [Бородкин, 2016].

«Историк с самого начала попадает в странное положение, – рассуждал Ю. М. Лотман, – в других науках исследователь начинает с фактов, историк получает факты как итог определенного анализа, а не в качестве его исходной точки» [Лотман, 1994, с. 354]. Таким же образом обстоит ситуация и при сборе исторических данных. В отличие от многочисленных современных датасетов социальных и гуманитарных наук, собранных иногда автоматическим, а иногда коллективным путем, после чего собранные датасеты многократно используются для научных исследований, в истории данные, базы данных собираются обычно для конкретного исследования самим автором или небольшим авторским коллективом, что уже не первое десятилетие вызывает дискуссии вокруг повторного использования таких данных (подробнее см. [Володин, 2023]). Таким образом, создание датасета оказывается не только технологическим, но и методологическим решением, когда необходимо сформулировать основные свойства изучаемого

явления, зафиксированные в исторических источниках, чтобы собрать их в цельную информационную модель [Цифровые гуманитарные исследования, 2023].

М. Таллер в одной из своих работ пишет: «Более формально, данные – это отметки в некоторой репрезентативной системе, которые можно хранить. Информация возникает, когда эти отметки помещаются в некоторый контекст. Так, “22°” – это данные. “Температура в этой комнате 22°” – это информация. Знания возникают, когда эта информация сталкивается со способностью извлекать из нее советы для действий» [Thaller, 2019, с. 161]. Историческое исследование опирается на исторические источники, из которых получает данные, которые необходимо связать для получения информации, и интерпретировать для получения знаний.

Соотношение реальности и баз знаний описывается по-разному. Весьма интересным подходом к этому вопросу является концептуальная эталонная модель CIDOC CRM, описывающую формальную онтологию, являющуюся теоретическим и практическим инструментом представления знаний и данных в области культурного наследия [Концептуальная эталонная модель..., 2024]. Никола Гуарино, один из разработчиков модели CIDOC, определяет формальную онтологию как спецификацию набора именованных понятий, используемых для описания и аппроксимации части реальности, к которой добавляется логическая теория первого порядка, сужающая предполагаемое значение названных понятий [Guarino, 1998]. Материальная реальность с точки зрения онтологического описания рассматривается как все, что обладает субстанцией, которая может быть воспринята с помощью органов чувств или инструментов. Материальная реальность ограничена пространством и временем. То, что происходит в нашем сознании или создается нашим разумом, также рассматривается как часть материальной реальности, поскольку это становится очевидным для других людей, по крайней мере по нашим высказываниям, поведению и результатам деятельности. Смысл создания онтологии состоит в том, чтобы определить единицы описания, т.е. сущности и отношения, которые мы можем зафиксировать и на которые мы можем сослаться, чтобы различать части реальности [Концептуальная эталонная модель..., 2024, с. 37–38]. Получается, что формальные наблюдения, которые становятся основой современных баз данных и датасетов, являются результатом абстракции свойств наблюдаемого целостного исторического источника.

Цифровой поворот в мировой историографии ускорил производство таких абстракций, которыми как раз и оперируют разнообразные исторические датасеты, и описывают многочисленные стандарты описания данных и метаданных (включая уже рассмотренный пример CIDOC, от классических MARC и Дублинского ядра до Schema и WikiData). Причем в случае исторического исследования опосредованные наблюдения, полученные из исторического источника, разделяются на метаданные (в классическом источниковедении часто они описываются как внешняя критика) и собственно данные (сведения, полученные путем внутренней критики).

При этом следует учитывать различие источнико-ориентированного и проблемно-ориентированного подходов в исторических исследованиях. Если в первом случае речь идет и максимально полене переносе информации исторического источника в базу данных (иногда такой результат даже описывается как создание самостоятельного вторичного источника), то во втором случае предполагается создание базы данных как средства систематизации исследовательского материала для последующих запросов и обработки формализованных наблюдений (база данных становится инструментом для решения конкретно-исторической проблемы, встречается понятие «историко-ориентированные информационные системы» [Корниенко и др., 2021]). Концепция источнико-ориентированного подхода относится к рубежу 1980–1990-х гг., когда интерес к теории и практике создания баз данных на основе материалов исторических источников был высоким. Автором этого подхода стал М. Таллер, который в 1987 г. сформулировал тезис о фундаментальном различии между обработкой информации исторических источников и обработкой данных в других науках, в том числе между источнико-ориентированным и проблемно-ориентированным подходами или способами использования компьютера в историческом исследовании [Таллер, 1993]. Таллер также доказывал, что опыт квантификации исторических наблюдений показал ограниченную применимость статистики к решению проблемы «размытых» исторических данных. Например, в предисловии к инструкции по использованию

созданной для таких целей системе управления базами данных Клею́ уточнялось, что «исторические науки, ввиду специфики данных, с которыми они главным образом работают, представляют такие требования к гибкости структуры данных, которые не могут быть выполнены в короткие сроки коммерческими программами», «с точки зрения дальнейшего развития науки скаживаются к тому же возрастающие особенности исторических данных (общая неотчетливость информации исторических источников, их сильная чувствительность к контексту и т.д.), решение которых также не приходится ожидать долгое время ни от кого, кроме заинтересованных в этих проблемах исторических наук» [Таллер, 1995, с. III]. Сегодня программное обеспечение пошло по пути применения общих как коммерческих, так и открытых решений. Тем не менее задачи оптимизации имеющихся программ для нужд исторических исследований стоят весьма остро (в том числе при создании специализированных больших языковых моделей, учитывающих исторический контекст и способных к алгоритмической проверке фактов).

Таким образом, при источнико-ориентированном подходе в исторических исследованиях необходимо максимально полно сформулировать весь перечень наблюдений как внешней, так и внутренней критики исторического источника, чтобы постараться не упустить ничего важного (как показывает практика, это фактически невозможно, и в дальнейшем, на следующем шаге вслед за пионерскими исследованиями ученые начинают обращать внимание на упущения прошлых наблюдений). При проблемно-ориентированном подходе в исторических исследованиях необходимо сформулировать исследовательские вопросы, которые уже будут складываться с систему наблюдений, заносимых в соответствующие информационные поля.

Основные постулаты применения технологий баз данных в исторических исследованиях появились еще в середине 1990-х гг. Ключевым исследованием в этой области была монография И. М. Гарской «Базы и банки данных в исторических исследованиях» [Гарская, 1994]. К тому же времени относится и начало жаркой дискуссии, к которой в размышлениях о формализации исторических наблюдений мы возвращаемся и 30 лет спустя. Речь о докладе нидерландского исследователя П. Доорна «Я и моя база данных: движение к концу направления “История и компьютеринг”», прозвучавшем в 1994 г. [Доорн, 1995]. Доорн сформулировал следующие проблемы компьютеризации исторической науки: обратил внимание на преобладание технической работы над аналитической, упрекнул в низком уровне методологической сложности работы с данными, отметил маргинализацию компьютеризованных результатов в академическом сообществе, указал на неудачу источнико-ориентированного подхода с акцентом на создание баз данных как самоцели, сделал прогноз о необходимости переориентации на междисциплинарные подходы и поиск целей с научно значимыми результатами. Критика Доорна сыграла роль катализатора: современная историческая информатика активно сочетает технические возможности с глубоким содержательным анализом исторических источников. Подробно содержание дискуссии и ее роль в развитии исторической информатики представлены в фундаментальной монографии «Историческая информатика. Эволюция междисциплинарного направления» [Гарская, 2018].

М. М. Кром приводит находчивое различие метода от подхода, использующегося в историческом исследовании [Кром, 2015, с. 127]. Так, метод легко описывается как последовательность операций и воспроизводится, сохраняет безличность и анонимность, свободен от какой-либо идеологии или мировоззрения, существует длительное время, слабо подвержен научной моде, не связан с каким-то одним научным направлением. В таком смысле формализация при создании баз данных оказывается устойчивым методом абстракции свойств от объектов реальности, в том числе и исторических источников. Однако сложности формализации исторических (шире – гуманитарных) наблюдений остаются на повестке дня. В качестве примера можно привести иллюстрацию методологического «моста» из монографии *Modelling Between Digital and Humanities* [Ciula et al., 2023, р. 2], который символизирует взаимодействие между цифровыми методами и гуманитарными науками, подчеркивая их принципиальные различия и необходимость методической интеграции. Авторы отмечают ключевые противоположности между цифровым (digital) представлением знаний – формальным, структурированным, основанным на правилах – и гуманитарным (humanities) – аналоговым, непрерывным, нюансированным, герменевтическим. Взаимодействие этих двух подходов как раз создает методологиче-

ское напряжение и отражает существующие эпистемологические противоречия. Метафора моста соединяет два «берега», символизируя моделирование как инструмент преодоления разрыва между гуманитарным и цифровым, подчеркивается модульность и гибкость используемых сейчас моделей, которые могут сочетать структуру (цифровое) с творчеством (гуманитарное). При этом мост стоит на «изменяющейся почве», что указывает на нестабильность и динамичность взаимодействия между разными гуманитарными дисциплинами, потому что сегодня вполне заметны сложности в создании устойчивых методологических связей из-за различий в методах и целях наблюдений. Основная идея авторов достаточно проста: моделирование выступает как посредник, позволяющий перекинуть мост между точностью цифровых инструментов и интерпретативной глубиной гуманитарных наук. Однако этот процесс требует баланса между жесткими, строгими структурами и гибкостью, иногда противоречивостью, которая порождает творческое напряжение. Такая метафорическая иллюстрация взаимодействия призывает к осознанному диалогу, где различия становятся ресурсом, а не барьером для инноваций.

Рассуждая о герменевтике как искусстве правильного понимания чужой речи Ф. Шлейермакер справедливо отмечал, что «работать на понимание без осознавания (правил) и прибегать к ним только в единичных случаях значит действовать непоследовательно» [Шлейермакер, 1994, с. 42]. Именно он сформулировал идею герменевтического круга как принципа толкования, основанного на диалектике части и целого. При этом важной составляющей этого круговорота смыслов является именно человек с его видением или точкой зрения. «Неизбежное движение по кругу, – как писал Г.-Г. Гадамер, – именно в том и состоит, что за попыткой прочесть и намерением понять нечто “вот тут” написанное “стоят” собственные наши глаза (и собственные наши мысли), коими мы это “вот” видим» [Гадамер, 1991, с. 18].

Как замечают коллеги из Центра цифровой истории в Университете Люксембурга, сегодня происходит «лабораторный поворот», когда от привычных научных центров, благодаря успехам технологий коммуникаций, цифровые гуманитарные исследования оказываются основой для создания концептуальных лабораторий, основанных на новых формах мышления, коммуникации и обучения [Fickers, Tatarinov, 2022, р. 1–22]. Появляются своеобразные «зоны обмена» (*trading zone*) как пространство взаимодействия между разными эпистемическими культурами, где встречаются историки, специалисты в информационных технологиях, филологи, философы. Такая концепция происходит из рассуждений П. Галисона о координации убеждений и действий [Галисон, 2004]. Смысл таких действий – в преодолении дисциплинарных барьеров через общий язык и договоренности о принципах междисциплинарности.

В таком варианте пересекающихся «зон обмена» цифровая герменевтика рассматривает такие проблемы, как критическая рефлексия над влиянием цифровых инструментов на исторические исследования, а значит, включает анализ алгоритмов, интерфейсов, визуализации данных и их роли в интерпретации прошлого. С практической точки зрения актуальным становится сочетание цифровой грамотности и критического мышления. Несмотря на успехи исторической информатики в России и цифровой истории в мире, цифровые технологии еще можно считать зоной эксперимента в исторических исследованиях (здесь идет речь и о компьютеризированном анализе текстов, и о географических информационных системах, и о примерах компьютерного моделирования). Но размышления о методе сегодня концентрируются над процессами перехода от «источников» к «данным» и затем обратно. Такое развитие требует коллaborации между дисциплинами с разными эпистемическими традициями, когда снимается конфликт между количественными и качественными методами, а смешанные методы позволяют подбирать необходимую сборку методов под конкретную историческую задачу. Также необходимым условием развития можно считать преодоление терминологических различий, например, между базовым для исследователей историков понятием «источники» и базовым для специалистов в области информационных технологий понятием «данные».

Цифровая герменевтика – это не только теория, но и практика *thinkering* (сочетание мышления и экспериментов), требующая постоянной рефлексии над тем, как технологии формируют наше понимание прошлого [Fickers, van der Heijden, 2020]. Таким образом, возникают разные стратегии критики «цифрового» в самом широком смысле этого слова. Внимание кри-

тического подхода направлено прежде всего на преодоление иллюзии «объективности» цифровых методов. Иллюзия «объективности» цифровых методов заключается в ошибочном восприятии технологий как нейтральных и беспристрастных инструментов, тогда как на деле они подвержены влиянию человеческих предубеждений, методологических ограничений и контекстуальных искажений. Такие убеждения можно назвать мифом о «технической нейтральности». Цифровые методы часто воспринимаются как «объективные» из-за их технической природы, хотя они встроены в социальные и культурные контексты (например, см. [Дастон, Галисон, 2018]). Так, использование больших данных в истории может создавать иллюзию «истины в числах», при этом игнорируются качественные аспекты развития. Цифровая герменевтика требует рефлексии над тем, как именно используемые инструменты (пере)формулируют исследовательские вопросы.

Перечень проблем, вокруг которых происходит дискуссия в цифровой герменевтике, достаточно широк: это и предвзятость алгоритмов, и селективность данных, и влияние инструментов и интерфейсов, и, наконец, упрощение сложности реальных наблюдений и связей [Rockwell, Sinclair, 2016]. Алгоритмы создаются людьми и обучаются на данных, которые не просто могут, а отражают исторические или социальные предрассудки. Например, NLP-модели (для обработки естественного языка), обученные на текстах с социальными стереотипами, будут воспроизводить их в результатах анализа. Поэтому критика алгоритмов направлена на выявление их «скрытой логики», которая может искажать интерпретацию данных, особенно в случае, когда алгоритм считается «черным ящиком» [Пасквинелли, 2024].

Цифровые инструменты (СУБД, ГИС, 3D) и интерфейсы (поисковые системы, архивы) задают рамки поиска и анализа, акцентируя одни аспекты и скрывая другие. Ранжирование результатов поиска создает иллюзию «важности» первых позиций, хотя это всего лишь результат алгоритмических решений. Получается, что интерфейсы цифровых архивов формируют восприятие информации через принципы навигации и визуализации. Причем понятно, что визуализации (как карты, так и графики) вынужденно упрощают многомерные исторические процессы до линейных моделей, что создает иногда ложное впечатление ясности. Так, карта торговых путей вынужденно игнорирует социальные или культурные факторы, сводя историю к географии передвижений. В этой связи крайне важна критика интерфейсов, указывающая на то, как дизайн визуализаций влияет на интерпретацию.

Иллюзия объективности может приводить к некритическому принятию результатов цифрового анализа, а также воспроизведению системных предрассудков, упрощению сложных гуманитарных проблем. Цифровые инструменты не заменяют интерпретацию, они становятся ее соавторами, требующими постоянной рефлексии и оценки промежуточных результатов анализа. Таким образом, преодоление этой иллюзии требует критического подхода (глубокого анализа алгоритмов, исторических источников, цифровых инструментов) и осознания, что цифровые методы и технологии не могут дать «окончательные ответы», а являются частью сложного и противоречивого диалога между технологическими возможностями и историческим знанием.

Одним из путей развития цифровой герменевтики является мультимодальная грамотность – способность понимать, интерпретировать и создавать информацию, используя различные формы представления данных (модальности), такие как текст, изображения, звук, видео, карты, 3D-модели и интерактивные интерфейсы. В контексте цифровой истории она подразумевает умение работать с разнообразными историческими источниками и инструментами, критически анализируя, как разные медиа влияют на восприятие и интерпретацию информации. Мультимодальная грамотность – это не просто технический навык, а критическое мышление, позволяющее историкам адаптироваться к цифровой эпохе, работать с разнообразными источниками и создавать многомерные интерпретации прошлого. Как образно отмечают коллеги, мультимодальная грамотность становится ключевым элементом «цифровой кухни» исследователя, где «сырые данные» превращаются в «приготовленные знания» через эксперименты и рефлексию [Fickers, Tatarinov, 2022, р. 16].

Сегодня данные часто «готовят» для исследователей специальные алгоритмы, к которым относят и большие языковые модели (LLM), и генеративные предобученные трансформеры

(GPT). Речь уже идет не только о применении известных сервисов (GPT, DeepSeek, GigaChat, YandexGPT и др.), сколько о построении таких моделей под собственные исследовательские нужды. Например, в книге Себастьяна Рашка *Build a Large Language Model From Scratch* представлено пошаговое руководство по созданию GPT-подобных больших языковых моделей (LLM) с нуля [Raschka, 2024]. Автор делает акцент на понимание архитектуры трансформеров и их практическую реализацию, последовательно разбирает этапы разработки LLM – от подготовки данных (токенизация через BPE, создание эмбеддингов, позиционное кодирование) и реализации механизмов внимания (self-attention, каузальные маски, multi-head attention) до сборки полноценной модели с трансформерными блоками, остаточными связями и слоевой нормализацией. Особое внимание уделяется предобучению на больших текстовых корпусах через задачу предсказания следующего токена, а также тонкой настройке под конкретные задачи, такие как классификация наблюдений или генерация инструкций. Книга сочетает теорию с практикой, предлагая примеры кода на PyTorch, упражнения и советы по оптимизации вычислительных ресурсов, что делает ее ценным ресурсом для разработчиков и исследователей, желающих глубоко понять внутреннюю работу современных LLM и создать свою модель, начиная с базовых компонентов и заканчивая интеграцией предобученных весов, как в известных коммерческих моделях GPT-2/3.

Важным направлением в улучшении качества результатов машинного обучения сегодня является построение исследовательских процессов на основе генерации, дополненной поиском (так называемый RAG-подход, *Retrieval-Augmented Generation*). Такой метод позволяет моделям генеративного искусственного интеллекта извлекать из баз данных или корпусов текстов информацию и включать извлеченную информацию в свои расчеты. Он изменяет принципы взаимодействия с большой языковой моделью (LLM), потому что она начинает отвечать на запросы пользователя уже с прямой ссылкой на указанный набор документов, используя эту информацию для дополнения информации из своих уже обработанных данных обучения. Это позволяет языковой модели использовать данные из конкретного объема исторических источников. Варианты такого использования включают предоставление, например, чат-боту доступа к собранным исследователем данным, а вывод будет основываться на отобранных им массивах данных (хотя и такой подход полностью не гарантирует отсутствия «галлюцинаций» у модели; под «галлюцинациями» обычно понимают такую выдачу модели, которая не имеет смысла или не соответствует действительности). Рассмотрим этот подход подробнее.

Технология генерации, дополненная поисковой информацией (RAG), может стать мощным инструментом в руках историка, работающего с методами машинного обучения, решительно преодолевая ключевые ограничения «чистых» генеративных моделей (генеративных предобученных трансформеров и больших языковых моделей) и усиливая их сильные стороны в контексте исторических исследований. Перечислим отличительные черты RAG-подхода.

Во-первых, он позволяет успешно бороться с так называемыми «галлюцинациями» языковых моделей, а вместе с этим дает шанс повысить фактическую точность выдачи чат-ботов. Стандартные большие языковые модели, обученные на обширных, но часто неспециализированных и потенциально содержащих многочисленные ошибки данных, склонны «галлюцинировать», т.е. генерировать правдоподобно звучащие, но фактически неверные утверждения (в исторических запросах такой материал встречается часто – условные даты, неточные имена, недостоверные события, выдуманные монографии, ошибочные интерпретации). Для историка такие действия алгоритма часто означают одно: таким алгоритмом пользоваться нельзя. RAG-подход может быть решением в таком случае, потому что перед генерацией ответа система извлекает данные из проверенных исторических источников (из подготовленного датасета). И генеративная модель основывает свой вывод строго на этой извлеченной информации. Эта возможность важнее, чем может показаться, потому что, по сути, такой подход позволяет обучить модель на данных, отражающих конкретный исторический период, тогда как большие языковые модели обычно лишены темпоральности и склонны к методологическому презентизму, так как основаны на современных интернет-источниках. В результате значительно снижается риск выдумывания фактов, ответы становятся надежно «заземленными» в реальных исторических источниках и научной литературе. Конечно, для этого необходимо тщательно подходить к отбору обучающих материалов.

Во-вторых, RAG-подход обеспечивает большую прозрачность результатов и повышает возможности для верификации полученных ответов. Большие языковые модели – это «черный ящик», когда историк не может понять, на основании каких именно данных модель пришла к тому или иному выводу, что критически важно для научной строгости. (Заметим, что существуют исключения: например, модель Perplexity Pro, позволяющая получить доступ к источникам генерации.) Ключевая особенность RAG-подхода – возможность показать источники, использованные для генерации ответа. Система может предоставить историку конкретные фрагменты текста из оцифрованных архивов, датасетов, статей, на которых основан ответ, приводятся ссылки на документы или публикации. Таким образом, историк получает возможность проверить обоснованность вывода, свериться с первоисточником, оценить контекст извлеченной информации, что критически важно для методологии и построения аргументации.

В-третьих, RAG-подход позволяет работать со специализированными данными. Исторические исследования часто фокусируются на узких темах, редких событиях, специфических коллекциях документов, просто не отраженных в общем обучающем корпусе больших языковых моделей. По этой причине традиционные модели плохо справляются с такими запросами. Историк может настроить базу знаний для поиска под свою конкретную задачу, чтобы решить эту задачу. Таким датасетом может быть любой подготовленный набор данных, например, оцифрованный личный архив изучаемой персоны, или коллекция писем или дневников определенного периода или региона, специализированная база данных по экономической истории, корпус научных работ по конкретной историографической проблеме. В итоге модель получает доступ к узкоспециализированной, актуальной информации, которой нет (и часто не могло быть) в ее общей памяти модели, позволяя историку задавать сложные, контекстно-зависимые вопросы по собственному уникальному материалу.

Запросы историка часто требуют сложного анализа в конкретном информационном контексте. Простые ответы больших языковых моделей могут быть поверхностными и упускать важные нюансы. Извлекая релевантные фрагменты из отобранных исторических источников, RAG-подход предоставляет генеративной модели богатый, специфический контекст. Модель не полагается только на свои обобщенные знания, а синтезирует ответ, глубоко интегрируя извлеченную информацию. Следовательно, генерируемые выводы или обобщенные наблюдения становятся более глубокими, нюансированными и релевантными конкретному историческому контексту, поставленной задаче и используемым источникам. Например, RAG-подход может помочь сравнить точки зрения разных историков на событие, извлекая соответствующие фрагменты из их работ.

С эвристической точки зрения ручной поиск информации в больших оцифрованных архивах весьма трудоемок, а формулировка сложных поисковых запросов для традиционных систем может быть затруднительна. Возможное подспорье – если историк сможет вести диалог с RAG-системой. В таком диалоге человека с машиной историк сможет задавать уточняющие вопросы по извлеченной информации, просить сравнить информацию из разных извлеченных источников, запрашивать резюме группы релевантных документов по теме на основе найденных данных. Получается, что RAG-система выступает в роли интеллектуального ассистента, позволяющего ему интерактивно и эффективно исследовать массив отобранных исторических источников, формулируя запросы на естественном языке и получая осмысленные, основанные на данных ответы, что ускоряет и обогащает процесс анализа.

Важно подчеркнуть, что RAG-система – это не замена исследователя, а удобный усилитель его возможностей. Как в свое время электронные таблицы или системы управления базами данных существенно усилили возможности исследователей в расчетах и выявлении взаимосвязей и поиске скрытых взаимосвязей, так и RAG-системы позволяют использовать новые и сильные возможности генеративных предобученных трансформеров, но с опорой на комплексы исторических данных с возможностью избежать классических «галлюцинаций» больших языковых моделей. Таким образом, исследователь сможет больше доверять выводам искусственного интеллекта благодаря привязке к источникам и возможностям верификации сгенерированного контента, фокусироваться на уникальных, специализированных коллекциях данных, а

вместе с тем и понимать, почему система выдала тот или иной ответ. RAG-подход позволяет углублять исторический анализ за счет интеграции релевантного контекста из исторических источников. Он помогает преодолеть главные опасения историков при работе с искусственным интеллектом – неточность ответов, «черный ящик» алгоритмов, оторванность от исходных источников. Данный подход позволяет раскрыть истинный потенциал машинного обучения для анализа исторических данных, делая взаимодействие с искусственным интеллектом безопасным, прозрачным и научно обоснованным.

Тем не менее важно подчеркнуть незаменимость историка-исследователя. Внедрение RAG-подхода (как и любого другого ИИ-инструмента) в историческую науку принципиально не заменяет историка, а усиливает его возможности, становясь интеллектуальным ассистентом исследователя. RAG-система может отлично находить, например, релевантные фрагменты текста по запросу к огромной коллекции исторических источников. Но понимание смысла этих фрагментов, их культурного, социального, политического и временного контекста пока лежит за пределами возможностей искусственного интеллекта. Историк интерпретирует нюансы языка, подтекст, иронию, пропаганду, личные мотивы автора источника – все то, что требует глубокого понимания человеческой природы и исторической эпохи. RAG-подход предоставляет «кирпичики» (например, текстовые фрагменты), но здание смысла (нarrатив и аргументацию) строит историк.

Самая ценная часть научной работы – формулировка исследовательских вопросов. Какие проблемы важны? Какие гипотезы стоит проверить? Какую точку зрения стоит использовать? Все это требует научной интуиции, знания историографии, понимания белых пятен в науке и творческого подхода. RAG-системы реагируют на запросы пользователя. Это мощный инструмент для поиска ответов, но не для формулировки значимых вопросов. Инициатива в постановке значимых, новаторских, методологически обоснованных вопросов целиком принадлежит историку.

RAG-подход и лежащие в его основе методы машинного обучения могут выявлять статистические закономерности, корреляции, кластеры. Но объяснение, почему эти паттерны возникли, каковы их исторические причины и долгосрочные последствия, в чем их культурное или социальное значение, – это сфера исторической интерпретации. Историк выходит за рамки констатации факта к объяснению. RAG-системы помогают увидеть паттерны, а историк может их объяснить. RAG-подход эффективен только в рамках той базы знаний (оцифрованных исторических источников, научной литературы), на основе которой он работает. Критическая оценка репрезентативности, полноты и системных ошибок самой этой базы знаний – задача историка. Он понимает, почему одни документы сохранились, а другие нет, чьи голоса представлены в архивах, а чьи замалчиваются, осознает эпистемологические ограничения доступного корпуса данных. RAG не может самостоятельно оценить границы и предвзятость источника.

RAG может суммировать информацию, компилировать данные из разных источников, но создание нового исторического знания, построение убедительного, доказательного и осмыслинного исторического нарратива, вписывающегося в контекст современной историографии, – это интеллектуальный акт творчества и синтеза, доступный только человеку-исследователю. RAG – инструмент в его руках, но не соавтор научного открытия или монографии. Историк берет на себя ответственность за свои выводы и интерпретации.

Образно говоря, RAG-система – это «цифровая лупа» современного историка. Она берет на себя трудоемкие задачи поиска, первичной обработки и структурирования огромных массивов информации, предоставляя отфильтрованный, релевантный запросу и верифицируемый материал. Такой инструмент освобождает драгоценное время и когнитивные ресурсы исследователя для главного – критического анализа, глубокой интерпретации, формирования суждений, построения аргументации и создания нового исторического знания. RAG-подход расширяет масштаб и эффективность работы историка, но не заменяет его экспертизу, критическое мышление, творческую интерпретацию и научную ответственность. Он становится помощником, ассистентом и учителем человеческого интеллекта в эпоху цифровой истории.

Цифровая герменевтика, вооруженная современными технологическими решениями, позволяет историку обращаться ко все большему числу исторических источников. Путь от историче-

ского источника через формализацию данных к новому знанию напоминает работу реставратора. Цифровые инструменты, подобно современным микроскопам и спектрографам, дают беспрецедентную детализацию, выявляют скрытые слои и связи, позволяют «очистить» источники от наслойений времени. Технология генерации, дополненная поисковой информацией, действует как интеллектуальный пинцет, бережно извлекающий нужные фрагменты мозаики прошлого из гигантского массива. Но сам акт интерпретации – восстановление целостной картины, понимание замысла и, главное, осмысление восстановленного – остается исключительной прерогативой историка, реставратора прошлого в данных. Цифровая герменевтика предоставляет инструменты невиданной точности, но не отменяет необходимости человеческого взгляда, способного отличить подлинный шедевр от искусственной подделки времени или алгоритма.

### Библиографический список

- Бородкин Л. И.* Моделирование исторических процессов: от реконструкции реальности к анализу альтернатив. СПб.: Алетейя, 2016. 306 с.
- Володин А. Ю.* Исторические исследования в контексте датаизма: методологический аспект // Вестник Перм. ун-та. История. 2023. № 4 (63). С. 135–147. DOI: 10.17072/2219-3111-2023-4-135-147. EDN: GFSIOJ.
- Володин А. Ю.* Между data и capta: проблемы датафикации исторических исследований // Вестник Перм. ун-та. История. 2019. № 3 (46). С. 137–145. DOI: 10.17072/2219-3111-2019-3-137-145. EDN: GGZDFX.
- Гадамер Г.-Г.* Философские основания XX века // Актуальность прекрасного. М.: Искусство, 1991. С. 15–26.
- Галисон П.* Зона обмена: координация убеждений и действий // Вопросы истории естествознания и техники. 2004. № 1. С. 64–92.
- Гарскова И. М.* Базы и банки данных в исторических исследованиях. Геттинген: Konrad Pachnicke Max-Planck-Institut fur Geschichte, 1994. 215 с. EDN: RZYNSR.
- Гарскова И. М.* Историческая информатика: эволюция междисциплинарного направления. СПб.: Алетейя, 2018. 408 с.
- Дастон Л., Галисон П.* Объективность. М.: НЛО, 2018. 584 с. ISBN: 978-5-4448-0932-7. EDN: PIXKTY.
- Доорн П.* Я и моя база данных: движение к концу направления «История и компьютеринг» // Информационный бюллетень ассоциации «История и компьютер». 1995. № 13. С. 48–77.
- Ковальченко И. Д.* Методы исторического исследования. 2-е изд., доп. М.: Наука, 2003. 486 с. EDN: XRPQKHN.
- Концептуальная эталонная модель CIDOC CRM / пер. с англ. и ред. А. Б. Антопольский, Д.Ю. Гук. М.: ИНИОН РАН, 2024. 400 с.
- Корниенко С. И., Гагарина Д. А., Поврозник Н. Г.* Исторические информационные системы: теория и практика. М.: Изд. дом НИУ ВШЭ, 2021. DOI: 10.17323/978-5-7598-2321-6. ISBN: 978-5-7598-2321-6. EDN: XVKOQD.
- Кром М. М.* Введение в историческую компаративистику. СПб.: Изд-во Европ. ун-та в С.-Петербурге, 2015. 248 с. ISBN: 978-5-94380-189-1. EDN: YWOTRW.
- Лотман Ю. М.* Изъявление Господне или азартная игра? (Закономерное и случайное в историческом процессе) // Ю. М. Лотман и тартуско-московская семиотическая школа. М.: Гнозис, 1994. С. 3533–3563.
- Пасквинелли М.* Измерять и навязывать. Социальная история искусственного интеллекта. М.: Individuum, 2024. 352 с.
- Таллер М.* Что такое «источнико-ориентированная обработка данных»; что такое «историческая информатика» // История и компьютер: новые информационные технологии в исторических исследованиях и образовании. Геттинген, 1993. С. 5–18.
- Таллер М.* Клэйш 4. Система управления банком данных / пер. с нем. и подг. к изданию М.М. Леренман. М.: Изд-во Моск. гор. объединения архивов, 1995. 279 с.
- Цифровые гуманитарные исследования: коллективная монография. Красноярск: СФУ, 2023. 272 с. URL: <https://bi.k.sfu-kras.ru/shop/publication?id=BOOK1-ББК71/Ц%20752-494468>
- Шлейермахер Ф.Д.* Герменевтика. СПб.: Европейский Дом, 2004. 242 с.

- Шмидт С. О. Современные проблемы источниковедения // Шмидт С.О. Путь историка. Избранные труды по источниковедению и историографии. М.: Изд-во РГГУ, 1997. С. 21–63.*
- Юмашева Ю. Ю. Источниковедение информационной эпохи. М.: Директ-медиа, 2024. 512 с. DOI: 10.23681/712244. ISBN: 978-5-4499-4221-0. EDN: ELNGSA.*
- Ciula A., Eide Ø, Marras C., Sahle P. Modelling Between Digital and Humanities. Thinking in Practice. OpenBook Publishers, 2023. 226 p. DOI: 10.11647/obp.0369.05.*
- Drucker J. Humanities Approaches to Graphical Display // DHQ: Digital Humanities Quarterly. 2011. Vol. 5, no. 1. URL: <https://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html>*
- Fairfield P. History and Hermeneutics. of Elements in Historical Theory and Practice. Cambridge: Cambridge University Press, 2025. 64 p.*
- Guarino N. Formal Ontology in Information Systems: Proceedings of the 1st International Conference June 6–8, 1998, Trento, Italy. 1st edn. Amsterdam: IOS Press, 1998. P. 3–15.*
- Fickers A., Tatarinov J. Digital History and Hermeneutics: Between Theory and Practice. Berlin, Boston: De Gruyter Oldenbourg, 2022. 312 p. DOI: 10.1515/9783110723991.*
- Fickers A., van der Heijden T. Inside the Trading Zone: Thinkering in a Digital History Lab // Digital Humanities Quarterly. 2020. 14, no. 3. URL: <https://www.digitalhumanities.org/dhq/vol/14/3/000472/000472.html>*
- Lavin M. Why Digital Humanists Should Emphasize Situated Data over Capta // DHQ: Digital Humanities Quarterly. 2021. Vol. 15, no. 2. URL: <https://www.digitalhumanities.org/dhq/vol/15/2/000556/000556.html>*
- Raschka S. Build a Large Language Model (From Scratch). Manning, 2024. 368 p.*
- Rockwell G., Sinclair S. Hermeneutica: Computer-Assisted Interpretation in the Humanities. Cambridge, London: MIT, 2016. 256 p. DOI: 10.7551/mitpress/9522.001.0001.*
- Romele A. Digital Hermeneutics: Philosophical Investigations in New Media and Technologies. Routledge, 2020. 168 p. DOI: 10.4324/9780429331893.*
- Schöch C. Big? Smart? Clean? Messy? Data in the Humanities // Journal of Digital Humanities. 2013. No. 2(3). P. 2–13.*
- Thaller M. Information Technology, Information and History // Вестник Перм. ун-та. История. 2019. № 3 (46). С. 159–174. DOI: 10.17072/2219-3111-2019-3-159-174. EDN: SMAMZK.*

*Дата поступления рукописи в редакцию 29.05.2025*

## **DIGITAL HERMENEUTICS OF HISTORICAL SOURCE: FORMALIZATION AS INTERPRETATION**

**A. Yu. Volodin**

Moscow State University, Lomonosovskiy ave., 27-4, Moscow, 119192, Russia

Siberian Federal University, Svobodny ave., 82A, Krasnoyarsk, 660041, Russia

volodin@hist.msu.ru

ResearcherID: J-9603-2012

Scopus Author ID: 57219052529

SPIN: 9627-6687

The article is devoted to understanding the transformation of the methodology of historical research under the influence of the “digital turn”, with a focus on the key discipline – source studies. The author examines how the rapid development of artificial intelligence (AI) technologies, especially machine learning, poses new challenges and opens up opportunities for historians, requiring a rethinking of traditional practices of working with historical sources. The central problem of the article is the process of transition from a historical source to formalized data (for example, through digital hermeneutics) and back to interpretation and new historical knowledge. The author emphasizes that this process is not just a technical operation, but a deeply methodological act of abstraction (using the example of collected historical data, or *capta*), associated with the inevitable loss of nuances and requiring critical reflection. The article examines methodological dilemmas of historical research, such as the opposition between source-oriented (*ad fontes*, striving for maximum completeness) and problem-oriented (*ceteris paribus*, focus on selected parameters) approaches, their evolution in the digital age, and related discussions. The “illusion of objectivity” of digital methods is critically analyzed. The author argues that algorithms, digital tools (DBMS, GIS, NLP), and interfaces are not neutral: they carry biases, limit focus, simplify complexity, and sometimes even predetermine research questions, which gives rise to risks: algorithm bias, data selectivity, the influence of visualization and interface

design, and uncritical acceptance of AI results. As a promising solution that overcomes the key limitations of generative models (hallucinations, “black box”, isolation from sources), the article considers in detail the RAG (Retrieval-Augmented Generation) approach, showing its advantages for a historian. The main conclusion of the article emphasizes the indispensable role of the historian-researcher in the era of AI. RAG systems and similar tools act not as a replacement, but as powerful “digital magnifying glasses” or intelligent assistants, taking on the labor-intensive tasks of searching and primary data processing, freeing up the historian’s time for critical interpretation, assessing the reliability of sources, building arguments, formulating research questions, and creating new historical knowledge.

*Key words:* data, capta, hermeneutics, formalization, historical informatics, historical research methodology, digital humanities research, digital history, RAG.

#### References

- Antopol'sky, A. B., & Guk, D. Y. (Trans.). (2024). *Konseptual'naya etalonnaya model' CIDOC CRM* [Conceptual reference model CIDOC CRM]. INION RAN.
- Borodkin, L. I. (2016). *Modelirovaniye istoricheskikh protsessov: ot rekonstruktsii real'nosti k analizu al'ternativ* [Modeling of historical processes: From the reconstruction of reality to the analysis of alternatives]. Aleteyya.
- Cifrovye gumanitarnye issledovaniya* [Digital humanities research]. (2023). Siberian Federal University. <https://bik.sfu-kras.ru/shop/publication?id=BOOK1-ББК71/Ц%20752-494468>
- Ciula, A., Eide, Ø., Marras, C., & Sahle, P. (2023). *Modelling between digital and humanities: Thinking in practice*. Open Book Publishers.
- Daston, L., & Galison, P. (2018). *Ob "ekтивности"* [Objectivity]. Novoe literaturnoe obozrenie.
- Doorn, P. (1995). Me and my database: Towards the end of the "History and Computing" field. *Informatsionnyy byulleten' assotsiatsii «Istoriya i komp'yuter»*, 13, 48–77.
- Drucker, J. (2011). Humanities approaches to graphical display. *Digital Humanities Quarterly*, 5(1). <https://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html>
- Fairfield, P. (2025). *History and hermeneutics*. Cambridge University Press.
- Fickers, A., & Tatarinov, J. (2022). *Digital history and hermeneutics: Between theory and practice*. De Gruyter Oldenbourg. <https://doi.org/10.1515/9783110723991>
- Fickers, A., & van der Heijden, T. (2020). Inside the trading zone: Thinkering in a digital history lab. *Digital Humanities Quarterly*, 14(3). URL: <https://www.digitalhumanities.org/dhq/vol/14/3/000472/000472.html>
- Gadamer, H.-G. (1991). Philosophical foundations of the 20th century. In *Akтуал'nost' prekrasnogo* [The relevance of the beautiful] (pp. 15–26). Iskusstvo.
- Galison, P. (2004). Trading zone: Coordinating action and belief. *Voprosy istorii estestvoznaniya i tekhniki*, 1, 64–92.
- Garskova, I. M. (1994). *Bazy i banki dannykh v istoricheskikh issledovaniyakh* [Databases and databanks in historical research]. Max-Planck-Institut für Geschichte.
- Garskova, I. M. (2018). *Istoricheskaya informatika: evolyutsiya mezhdisciplinarnogo napravleniya* [Historical information science: Evolution of an interdisciplinary direction]. Aleteyya.
- Guarino, N. (1998). *Formal ontology in information systems: Proceedings of the 1st International Conference*. IOS Press.
- Kornienko, S. I., Gagarina, D. A., & Povroznik, N. G. (2021). *Istoricheskie informatsionnye sistemy: teoriya i praktika* [Historical information systems: Theory and practice]. Higher School of Economics Publishing House.
- Kovalchenko, I. D. (2003). *Metody istoricheskogo issledovaniya* [Methods of historical research] (2nd ed.). Nauka.
- Krom, M. M. (2015). *Vvedenie v istoricheskuyu komparativistiku* [Introduction to historical comparative studies]. European University at Saint Petersburg.
- Lavin, M. (2021). Why digital humanists should emphasize situated data over capta. *Digital Humanities Quarterly*, 15(2). <https://www.digitalhumanities.org/dhq/vol/15/2/000556/000556.html>
- Lotman, Y. M. (1994). The Lord's Proclamation or a Gamble? (The regular and the random in the historical process). In *Y.M. Lotman i tartusko-moskovskaya semioticheskaya shkola* [Y.M. Lotman and the Tartu-Moscow Semiotic School] (pp. 353–363). Gnozis.
- Pasquinelli, M. (2024). *Izmerit' i navyazyvat'. Sotsial'naya istoriya iskusstvennogo intellekta* [The eye of the master: A social history of artificial intelligence]. Individuum.
- Raschka, S. (2024). *Build a large language model (from scratch)*. Manning.
- Rockwell, G., & Sinclair, S. (2016). *Hermeneutica: Computer-assisted interpretation in the humanities*. MIT Press.
- Romele, A. (2020). *Digital hermeneutics: Philosophical investigations in new media and technologies*. Routledge.
- Schleiermacher, F. D. (2004). *Germenevtika* [Hermeneutics]. Evropeyskiy Dom.
- Schmidt, S. O. (1997). Modern problems of source studies. In *Put' istorika. Izbrannye trudy po istochnikovedeniyu i istoriografii* [The path of a historian: Selected works on source studies and historiography] (pp. 21–63). RGGU.

- Schöch, C. (2013). Big? Smart? Clean? Messy? Data in the humanities. *Journal of Digital Humanities*, 2(3), 2-13.
- Thaller, M. (1993). What is "source-oriented data processing"; what is "historical information science". In *Istoriya i komp'yuter: novye informatsionnye tekhnologii v istoricheskikh issledovaniyakh i obrazovanii* [History and computer: New information technologies in historical research and education] (pp. 5–18).
- Thaller, M. (1995). *Kλειό 4. Sistema upravleniya bankom dannykh* [Kλειό 4: Databank management system] (M. M. Lerenman, Trans.). Moscow City Archives Association.
- Thaller, M. (2019). Information technology, information and history. *Perm University Herald. History*, 3(46), 159–174.
- Volodin, A. Y. (2019). Between data and capta: Problems of datafication of historical research. *Perm University Herald. History*, 3(46), 137–145.
- Volodin, A. Y. (2023). Historical research in the context of dataism: Methodological aspect. *Perm University Herald. History*, 4(63), 135–147.
- Yumasheva, Y. Y. (2024). *Istochnikovedenie informatsionnoy epokhi* [Source studies of the information age]. Directmedia.