

ЦИФРОВАЯ ЭПИСТОМОЛОГИЯ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Цифровая эпистемология и информационные технологии

Лойко А.И.

Цифровая эпистемология играет важную роль в изучении проблем, связанных с общей теорией познания и когнитивными науками [1]. В числе этих проблем трансформация процессов получения и обработки данных в форматы информации и знаний [2]. В этой трансформации активную роль играют технологии искусственного интеллекта [3].

Они благодаря разработчикам сформировали обширную инфраструктуру, основанную на архитектуре цифровых экосистем. В предметном поле семантического искусственного интеллекта актуальной стала тема достоверности данных и информации, поскольку получили распространение практики фейковых технологий и цифрового мошенничества.

Подобные практики сделали востребованной тематику кибернетической безопасности. В её разработке активную роль играет философия права и цифровая этика. Цифровая эпистемология включает уровни гносеологии данных, информации и знаний. Они обладают спецификой, которая фиксируется категориальными структурами.

Тематика познания возникла, когда у людей сформировалась интенция на знания. Она была обусловлена практическими причинами. Одной из таких причин была деятельность в виде охоты, рыболовства и собирательства.

Охотнику нужно было знать поведение разных животных, их слабые и сильные стороны. При разделке туши тоже были нужны знания. Знания были нужны для изготовления орудий охоты и для организации охоты. Охотники должны были знать о животных практически всё, от чего зависела успешная охота. В таком же положении находились рыболовы и собиратели.

Каждый вид практической деятельности формировал интенцию на практические знания, включая земледелие и ремесло. Это были инструментальные

знания. Они обычно нигде не записывались. Они передавались устно из поколения в поколение. Когда люди начали оперировать значительными объёмами данных и информацией, то у них возникла необходимость в письменности и интеллектуальных технологиях вычисления и обозначения данных.

Для данных был выбран цифровой формат. Для информации был выбран формат письменного естественного языка. Информация в виде текста естественного языка содержала данные. Она отправлялась адресату и выполняла функции общения, в том числе, в деловом формате.

Поскольку деловые отношения требовали письменных правовых гарантий, то сформировалась практика юридических знаний в виде законов, указов, договоров. Это были нормативные знания, на основе которых строились деловые отношения между людьми и функционировали институты общества.

Философы реализовали интенцию значительной части человечества на концептуальные знания, которые используют язык категорий для создания общей онтологии. Эта онтология не систематизировала знания, а строилась каждым философом в собственной версии. При этом вскоре обнаружилось, что в общей онтологии есть общие категории для всех её авторских модификаций. Авторские версии общей онтологии множились по мере увеличения числа философов в этой тематике. Общая онтология давала ответ на вопрос, как устроен универсум.

Философы обнаружили для себя и такой предмет изучения как познавательные процессы в голове человека. Результатом стала разработка Аристотелем формальной логики. В ней он описал особенности реализации человеком мыслительных процессов в форме понятий, суждений и умозаключений. Им описаны модификации индуктивных и дедуктивных умозаключений.

Платон поднял важный вопрос о наличии у человека знаний до самого познавательного процесса. Он ответил на него утверждением, что знания у человека есть изначально. Ему остаётся их только вспомнить благодаря конкретному контексту.

Философы не стремились искать доказательства для создаваемых ими общих онтологий. Они ограничивались убеждением, что универсум устроен так, как они его конструируют в форме общей онтологии.

Ситуация в философии изменилась в новое время, когда стали формироваться структуры научного познания на теоретическом и эмпирическом уровнях познания. Для науки важными стали критерии ясности, точности, непротиворечивости, доказательности и обоснованности. Ф. Бэкон видел исследовательскую науку на основе эмпиризма и метода индукции. Он, как и Г. Галилей, важную роль отводил экспериментальным опытным исследованиям. Р. Декарт видел науку на основе рационализма и метода дедукции.

Он важную роль отводил логическим основаниям доказательств. С этой целью он разработал язык алгебры и сделал геометрию аналитической геометрией. Язык уравнений делал математику строгой наукой. Такой же критерий был применим и к механике (физике). Д. Юм придерживался позиции скептицизма. Она предписывала критически относиться к любым утверждениям. Ещё больший критицизм продемонстрировал И. Кант после того, как дискуссия о природе туманностей с его участием закончилась ничем. Немецкий философ сформулировал вопрос о возможности познания человеком природы такой, какой она есть вне пределов его сознания.

При этом в отличие от сенсуализма Локка Кант считает, что у человека в его мышлении (рассудке) есть априорные знания в виде категорий, которые используются для аналитики образов внешних объектов, являющихся сознанию человека. При этом в объектах есть непознаваемая для человека часть, называемая вещью-в-себе. Кант допускает и наличие апостериорных знаний.

Эмпиризм, сенсуализм и рационализм делали акцент на разную роль чувственных и рациональных форм познания. Эти дискуссии не мешали развитию наук о природе. Они не мешали развитию математики. Они не мешали формированию гуманитарных наук. Тематика гносеологии явно доминировала в структуре общей теории познания и была тесно связана с институтом науки. Представители марксизма и прагматизма одними из первых сделали

предметом изучения познавательную деятельность в связи с практикой. Представители прагматизма в аспект практики рассматривали особенности обыденного познания людей, связанного с опытом.

В тематике гносеологии в общем описании процессов познания важная роль отводилась исследованию познавательных процессов через изучение особенностей чувственных и рациональных форм познания. Поскольку процесс познания имеет результат в виде знаний, то актуальной стала тематика эпистемологии. В классической эпистемологии любой результат познания рассматривался как знания. Деления результатов познания на данные, информацию и знания не делалось. Любой результат познания рассматривался как знание. Только возникал вопрос, насколько оно является объективным и достоверным. Объективность знаний отстаивал Платон, который видел их основание в объективном идеализме.

Марксизм признавал объективность знаний исходя из позиции материализма, согласно которой знания являются результатом познания материальной реальности, существующей независимо от человека. Представители марксизма критиковали объективный идеализм за отрицание им первичности материальной реальности. Они также критиковали субъективный идеализм и агностицизм за отрицание ими возможности познания человеком внешней по отношению к нему объективной реальности.

Для агностицизма и субъективного идеализма знания сводятся к знанию того, как человек конструирует априорными структурами мышления явление внешней реальности его индивидуальному сознанию. В этом контексте возникла проблема достоверности знаний

Представители позитивизма, исходя из методологической платформы эмпиризма, стали применять ко всему массиву знаний критерий верификации, основанный на прямом соотношении атомарных предложений с наблюдаемыми фактами существования описываемых ими процессов, событий и объектов. В результате они обнаружили, что большая часть знаний, относящихся к

философии, является не верифицируемой. Также они обнаружили, что критерий верификации не применим к теоретическим знаниям ряда наук.

Представители неопозитивизма стали искать критерии достоверности их методологии в пространстве логики. Решением для них стала разработка аппарата математической логики. Эта логика оказалась востребованной для практических задач развития систем связи на основе цифровых технологий, использующих аппарат математической логики (двоичный код). Неопозитивизм стал одним из оснований для формирования аналитической философии.

Эта философия приняла активное участие в разработке математической логики, адаптированной к цифровым технологиям. Она стала частью информатики в форматах чётких и нечётких логик. Эти логики используются в программировании, в частности, при разработке больших языковых моделей.

Если классическая эпистемология больше ориентирована на обеспечение научного исследования стандартами достоверности, то цифровая эпистемология занята вопросом как компаниям и банкам эффективно использовать данные, информацию и знания для получения прибыли, и решения, связанных с ней задач. В таком ракурсе использования информационных ресурсов объективность и достоверность не играют основной роли. При этом под информационными ресурсами понимаются не все данные, информация и знания, а только та их часть, которая переведена в цифровой формат технологий искусственного интеллекта. Эти технологии стали прикладными инструментами цифровой рыночной экономики и политики.

Поскольку цифровые поколения пользуются данными, информацией и знаниями через мобильный телефон, интегрированный в социальные сети интернета, то для них это основной новостной ресурс, в который входят не только политические новости, но и реклама, навигация, новости через общение со сверстниками. Через формат цифровых технологий цифровые поколения получают доступ к научным и профессиональным знаниям.

Особенность цифровых технологий заключается в том, что они не только представляют пользователям социальные информационные ресурсы,

но и получают от них в режиме обратной связи на безвозмездной основе прирост этих ресурсов, поскольку пребывание пользователей в социальных сетях формирует приращение информационных ресурсов, что крайне выгодно компаниям, владеющим цифровыми платформами и цифровыми экосистемами.

Основными социальными информационными ресурсами для людей являются данные, информация и знания. С точки зрения науки данными являются сведения, полученные эмпирическими методами измерения, научного наблюдения и эксперимента, а также, методами формальной логики, в частности, математической логики, например, в виде арифметических действий.

В более узком смысле под данными понимаются сигналы, сведения и сообщения в форме цифрового языка совместимого с технологиями искусственного интеллекта для их хранения, обработки и использования.

В более широком онтологическом формате под информацией понимается объективное содержание сигнала, сообщения, сведений. Расшифровка этого объективного содержания входит в задачи современной науки. В этих целях используются автоматизированные системы сбора и обработки информации. Под социальной информацией понимаются сведения, которые люди передают друг другу. На рынке информационных услуг оперативной передачей сведений о событиях занимаются медиа холдинги.

Под социальной информацией также понимается значение, которое люди на бытовом и пользовательском уровне вкладывают в сигналы, сведения и сообщения. Тем самым они осуществляют интерпретацию данных исходя из собственного мировоззрения и понимания событий.

Реклама строится на максимальном приближении к мировоззрению пользователей. Ей свойственно «усиливать» эффект воздействия на сознание пользователей социальных сетей. Основной акцент делается на новизну информации. Частое повторение рекламы трансформирует её в информационный шум и вызывает раздражение у её пользователей.

Смещение пребывания современного человека из системы знаний в систему социальной информации отразилось на цифровых поколениях в форме

клипового мышления. Это мышление фрагментировано архитектурой сообщений. Оно теряет цельность семантического восприятия информационного поля. И в таком формате оно становится уязвимым для манипулирования им в рамках информационных войн.

Согласно определению, К. Шеннона, полезная функция информации заключается в том, что она сокращает степень неопределённости у человека в конкретных вопросах, связанных с его жизнедеятельностью.

Справочная социальная информация позволяет её пользователю уменьшить неопределённость при принятии решений. Сигнальная социальная информация позволяет её пользователю быть в курсе текущих событий и оперативно строить свою деятельность.

Коммерческая информация позволяет её пользователю отслеживать важные для него экономические показатели. Конфиденциальная информация носит закрытый характер для других пользователей, что обусловлено фактором кибернетической безопасности. Биографическая информация может быть открытой для других пользователей.

Социальная информация представляется в виде схем, эскизов и изображений. Она также представляется в виде графиков, диаграмм, символов, видео, фотографий и текста. Она представляется в виде презентаций, обзоров, образцов и является важной частью выставочной деятельности. Важную роль в подаче информации играет дизайн.

Социальная информация может иметь различные уровни допуска и иметь гриф секретности. Информация и данные, имеющие такой статус, являются предметом повышенного интереса. Этот интерес реализуется институтами разведки, контрразведки, промышленного шпионажа и хакинга.

Знания вводят данные и информацию в формат системного описания процессов, свойств и объектов с позиции общей, предметной и целевой онтологии, а также онтологии ценностей. Они конфигурируют данные и информацию в контекст научного объяснения, инженерного решения и практического действия. Знания вследствие этого содержат алгоритм решения типовых задач,

методы, модели, нормативные предписания. В настоящее время формат пользования знаниями ограничен образованием, профессиональной деятельностью и научными исследованиями.

В цифровой экономике знания как ресурс используются инжиниринговыми и консалтинговыми компаниями. Эти компании создают на основе знаний инновационный продукт в виде проекта и реализуют его под интересы заказчика с последующим обслуживанием созданной ими инфраструктуры и коммуникаций. Эти компании привлекают подрядные и субподрядные организации. Консалтинговые компании занимаются аналитикой, аудитом и бизнес планированием [4].

Есть концептуальный уровень знаний, который требует владения математическим аппаратом, специальными формализованными языками. Его архитектуру формируют законы, теории, принципы, постулаты, идеалы и нормы научной деятельности.

Есть эмпирический уровень знаний, архитектуру которого формируют факты, процедуры верификации и методы сбора и обработки данных. Когда данные, информация и знания переводятся в вычислительные системы, то это делается с помощью чётких и нечётких логик. В больших языковых моделях используются нечёткие логики. В этой логике вероятность члена множества определяет степень правомерности отнесения этого члена к множеству.

Чёткие и нечёткие логики используются и в функциях аналитики. Основным ресурсом для них являются формализованные структуры, пригодные для вычислений. Всё, что оцифровывается и попадает в систему, становится частью единого вычислительного мира. ИИ не знает, что он анализирует. Он обучается структурам данных.

Он фиксирует статистические связи между словами. Он не видит изображение. Он распознаёт паттерны пикселей. Он не слышит музыку. Он кодирует частотные распределения. Он формирует конфигурации данных в пространстве вероятностей. Он обнаруживает закономерность каузальных связей. Там, где человек интерпретирует, ИИ выявляет связи [5].

Чем больше цифровых данных, тем точнее функционирует большая языковая модель [6]. Чем разнообразнее спектр цифровых данных, тем богаче её латентное пространство. Чем качественнее цифровые данные, тем устойчивее реализуется обучение. Искусственный интеллект не создаёт новые смыслы. Он выявляет скрытые конфигурации структур данных [7].

Под данными понимаются сведения в формализованном виде, доступном для анализа. Это могут быть числа, тексты, изображения, звук, видеопотоки, телеметрия или логи работы систем. Данные существуют там, где что-то может быть прочитано ИИ. Они создают для ИИ возможность мышления в форме вычисления. 80% времени специалистов по данным тратится на процессы обеспечения качества данных. Эти процессы включают очистку, подготовку и валидацию данных [8].

Норберт Винер (Norbert Wiener) и Питер Друкер (Peter Drucker) указывали, что данные, информация и знание образуют разные уровни когнитивной структуры эпистемологии. Этот тезис используется в теории управления знаниями. Под данными (Data) понимаются необработанные наблюдения, фиксации и измерения. Смысл и контекст в них ещё не расшифрован [9].

Под информацией (Information) понимаются данные, помещённые в систему отношений, когда между элементами появляется структура интерпретации. В данном контексте возникают вопросы кибернетической безопасности. Они заключаются в защите процесса получения и хранения данных и информации. Особое внимание уделяется цифровым следам. Мониторинг заключается в выявлении слабых мест в цифровой среде и защите эпистемологического процесса от несанкционированного вмешательства.

Кибернетическая безопасность включает систему методов, технологий и практик, обеспечивающих защиту цифровых систем, социальных сетей, компьютерных программ и данных от атак, кражи, повреждения или несанкционированного доступа. Она осуществляет защиту личной информации, критической инфраструктуры и финансовых ресурсов, основываясь на принципах

конфиденциальности, целостности и доступности данных. В этих целях осуществляется использование антивирусов, сложных паролей и двухфакторной аутентификации.

Кибернетическая безопасность пресекает три вида угроз. Одной из них является кибернетическое преступление. Это действия, организованные одним или несколькими злоумышленниками с целью атаковать компьютерную систему, чтобы нарушить ее работу и извлечь финансовую пользу.

Методологией кибернетической атаки реализуются действия, имеющие целью манипулирование индивидуальным и общественным сознанием. Они являются частью информационной войны с пользователями социальных сетей. Кибернетический терроризм содержит действия, связанные с формированием страха у пользователей социальных сетей с целью вербовки их для осуществления террористической деятельности.

Вредоносное программное обеспечение является инструментом кибернетических преступников. Оно распространяется под видом файлов или почтовых вложений. Вирусы заражают файлы вредоносным кодом. Чтобы распространяться внутри системы компьютера, они копируют сами себя. Троянцы маскируются форматом лицензионного программного обеспечения. Шпионское программное обеспечение тайно следит за действиями пользователя социальной сети, и собирает информацию о нём, например, данные кредитных карт. Программы вымогатели шифруют файлы и данные пользователя социальной сети. Затем кибернетические преступники требуют выкуп за восстановление данных с оплатой.

Для распространения вредоносного программного обеспечения злоумышленниками используются также приложения рекламного характера. Злоумышленниками используются также ботнеты, SQL-инъекция и фишинг. В их арсенале находятся атаки Man-in-the-Middle и DoS-атаки.

Данные генерируются повсюду. Текст, изображение, движение, звук и метка GPS становятся частью потока данных, из которого формируются обучающие корпуса искусственного интеллекта. На текстовых данных книги,

статьи, сайта, форума, научного архива, переписки и технической документации обучаются языковые модели.

К визуальным данным относятся изображения и видео, собранные из открытых источников или сгенерированные системами компьютерного зрения. К звуковым и речевым данным относятся записи речи, подкасты, субтитры и аудиокниги, используемые для обучения систем распознавания речи и голосовых ассистентов. К сенсорным и телеметрическим данным относятся сигналы от датчиков, камер, навигационных систем, промышленных и медицинских приборов. К поведенческим данным пользователей относятся клики, просмотры, маршруты и время реакции. Эти данные фиксируют не содержание, а профиль действия пользователей социальных сетей. Они используются для рекомендательных систем и анализа предпочтений [10].

Всё превращается в данные процессом оцифровки. Текст, изображение, звук и движение стали единообразными с точки зрения вычислений. Любой элемент опыта можно выразить через числа. Слово трансформируется цифровой технологией в токен (единицу текста, преобразованную в индекс). Изображение трансформируется цифровой технологией в матрицу пикселей. Звук трансформируется цифровой технологией в спектрограмму частот. Движение трансформируется цифровой технологией в координаты и временные метки.

Ручная разметка стала невозможной. На смену ей пришли автоматизированные методы. Посредством краулинга (crawling) осуществляется автоматический сбор текстов из интернета. Сенсорные сети осуществляют сбор физических данных с устройств IoT (Internet of Things) и датчиков. API и телеметрия осуществляют сбор данных через программные интерфейсы, фиксирующие поведение систем и пользователей социальных сетей. Генерация синтетических данных имеет целью создание искусственных примеров на основе уже существующих наборов данных.

Искусственный интеллект участвует в сборе данных о самом себе. Модели создают синтетические тексты, изображения и сценарии, которые ис-

пользуются для обучения других интеллектуальных систем. Это форма рекурсивного цикла. ИИ порождает данные, на которых обучаются новые интеллектуальные системы. В результате формируется мета уровень данных.

ИИ оперирует только тем, что было в его данных. Большинство обучающих наборов данных создаются в США и Китае. Это приводит к тому, что глобальные модели содержат материалы медиа. В результате возникает вопрос, насколько данные отражают разнообразие опыта человечества. Если набор данных не включает все естественные языки, ИИ не сможет их понимать. Если в данных отсутствует опыт определённых культур, ИИ не сможет его реконструировать. Если данные искажены политическими или идеологическими фильтрами, искажённым будет и поведение модели.

Контроль над данными стал формой власти. Тот, кто решает, какие данные собирать, решает, каким будет знание. Для характеристик фотографии важным является дата съёмки, географическая локация и устройство съёмки. Для характеристик текста важным является автор, язык, время публикации и лицензия. Для характеристик набора данных важным является источник, объём, структура и способ аннотации.

Чтобы данные могли обучать искусственный интеллект, они должны быть не только собраны, но и объяснены. Аннотация (annotation) обозначает процесс добавления меток, категорий, тегов, которые связывают данные с понятиями. Без аннотации нейронная сеть не знает, что изображено или сказано. Она лишь оперирует числами. Аннотация делает возможным обучение с учителем (supervised learning), когда модели сопоставляют входные данные с правильными ответами и учатся по образцу.

Но аннотация несёт в себе субъективность. Разные люди по-разному видят одно и то же явление. Поэтому даже в машинном обучении человек остаётся источником смещения (bias) [11]. Но без аннотации невозможно было бы создание большинства современных ИИ-моделей от систем распознавания лиц до генеративных трансформеров.

Формат влияет на то, как данные читаются, интерпретируются и комбинируются. Он определяет границы памяти искусственного интеллекта. Он задаёт, что можно извлечь из данных. Поэтому архитектура данных играет ключевую роль. Но на начальном этапе работы с ними искусственного интеллекта они являются только потенциальным информационным ресурсом. В рабочее состояние этот ресурс приводит векторизация. Она осуществляет перевод данных в числовую форму, доступную для обработки нейронными сетями.

Векторы называются эмбедингами (embeddings). Они создают многомерную карту, где близкие по смыслу элементы располагаются рядом, а далёкие смыслы располагаются на расстоянии. Благодаря этому ИИ может сопоставлять, искать, прогнозировать и классифицировать. Всё это основывается на отношениях между векторами, а не на понимании.

Для ИИ смысл ассоциируется со структурой близостей в эмбединг-пространстве. В этом пространстве генерируется то, что люди называют знанием. После преобразования данных в векторы начинается обучение ИИ. Это процесс, в котором языковая модель ищет закономерности, связывая входные данные с внутренними параметрами. В классическом машинном обучении (machine learning) это происходит через оптимизацию весов числовых коэффициентов, регулирующих силу связей между нейронами.

Когда языковая модель делает ошибку, она корректирует веса с помощью алгоритма обратного распространения ошибки (backpropagation), разработанного Полом Вербосом.

То, что люди называют знанием, в машинной архитектуре существует как распределённая память. Это миллиарды весов, в которых зафиксированы статистические связи между элементами обучающего набора.

Искусственный интеллект полностью зависит от тех данных, на которых обучен. Это делает качество данных решающим фактором. Если данные неполные, шумные или смещённые, модель воспроизводит эти ошибки в своих выводах. Поэтому модели, обученные на ограниченных корпусах, демонстрируют культурные или языковые искажения. Когда языковая модель отвечает

на вопрос или создаёт текст, она активирует веса, изменённые в ходе обучения. Она воспроизводит закономерности, закреплённые в данных. Её ответ является реактивацией структурных следов.

В современных интеллектуальных системах эмбединги сохраняются отдельно от основной модели. Это позволяет моделям обращаться к внешним источникам знаний, находить похожие случаи и формировать отклики, более приближённые к контексту. В результате создаётся гибридная память. Часть её находится в весах, а ещё одна её часть находится во внешних данных.

В цифровой онтологии информацией является устойчивая конфигурация данных. Она не существует до акта обработки, но возникает в момент вычисления. Корпорации, пользующиеся искусственным интеллектом, не только используют данные. Они присваивают процесс наблюдения. Данные пользователей становятся сырьём для обучения моделей, часто без согласия пользователей социальных сетей. Этот феномен Шошана Зубофф (Shoshana Zuboff) назвала «капитализмом слежки» (*surveillance capitalism*). Это система, где личная жизнь становится источником прибыли.

Каждый запрос в поиске, каждая фотография, каждая пауза перед кликом становится элементом статистической модели, обучающей другие интеллектуальные системы. В результате сформировались угрозы приватности в эпоху искусственного интеллекта. Деанонимизацией обозначается возможность ИИ восстановить характеристики конкретной личности по обезличенным данным. Термином «Утечки данных» обозначается потеря контроля над хранилищами ИИ, где собирается личная информация.

Термином «Профилирование» обозначается построение психологических и поведенческих моделей пользователей для рекламы, политики или манипуляций. Термином «Инференция» обозначается способность ИИ обнаружить конфиденциальную информацию, даже если она явно не предоставлена в социальной сети.

Данные производят смысл. Мышление становится производной статистики. Каждый объект оставляет цифровой след. Каждое действие оставляет

координату. Каждое состояние оставляет сигнал. В технологиях ИИ произошёл переход к навигации в огромных объемах цифровых данных. В результате цифровая среда стала влиять на мышление человека. Особенно это заметно по роли цифровых платформ и цифровых экосистем в организации информационного пространства современного человека.

Информация создается цифровым контекстом, поскольку он предопределяет специфику эпистемологического объекта. Цифровые данные считаются объективным доказательством (*objective evidence*).

Содержащиеся в тексте языковые единицы и их сочетания, и выявленный цифровой алгоритмом паттерн становится нередко самодостаточным свидетельством значения. Но за нейтральными и объективными корпусами стоят интерпретаторы дизайнеры аналитических параметров их формирования. Цифровые технологии обнаруживают закономерности, но не предлагает объяснений. Поэтому обычно вмешательство человека встроено в цифровой объект, если рассматривать его как историю контактов человека с технологическим приспособлением.

Человек как агент приспосабливается к технологической системе и вводит те данные, которые она может обработать. Технологическая система выдает ответы. Является важным не столько сам цифровой объект, сколько инфраструктурная система отношений вокруг него.

Обработка текста может давать ряд токенов на одно и то же неверно распознанное слово, что требует дополнительной правки вручную. В итоге выясняется, что токен как вид данных является не объективно данным в тексте феноменом, а конструктом, возникающим во взаимодействии человека и инструмента. Одно и то же слово может восприниматься в разных контекстах, но это будет один и тот же токен.

Игра слов, и обычное употребление будут распознаваться как одна единица. Между двумя токенами в тексте можно установить статистически значимую связь, но это не всегда может иметь смысл. Вычленять в тексте слова,

которые с наибольшей вероятностью будут встречаться рядом трудно, поскольку между ними есть совпадение, но не обязательно есть причинно-следственные отношения.

Разработка пользовательского интерфейса, делающего интервенции в корпус исходных данных возможными может быть решением проблемы. Процедуры отбора, селекции, обработки данных в этом интерфейсе прозрачны. Каждая интервенция предлагается пользователю отдельно, и ее возможные ограничения учитываются заранее. Цифровая эпистемология в области политической философии изучает тематику, связанную с понятиями цифрового империализма и цифрового колониализма [12].

Список используемых источников

1. Viola, L. *Humanities in the Digital: Beyond Critical Digital Humanities*. N.Y.; L.: Palgrave Macmillan, 2023. – XXVI, 173 p.
2. Carter, J. A. *Digital Knowledge A Philosophical Investigation*. – Routledge, 2025 – 206p.
3. Dotan, R. Theory choice, non-epistemic values, and machine learning // *Synthese*. 2021. Vol. 198. P. 11081 – 11101.
4. Grim, P., Singer, D. *Computational Philosophy* // *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition) / Ed. by E.N. Zalta, U. Nodelman. URL: <https://plato.stanford.edu/archives/fall2022/entries/computational-philosophy/>
5. Ingvarsson, J. *Towards a Digital Epistemology: Aesthetics and Modes of Thought in Early Modernity and the Present Age*. Cham, Switzerland: Springer International Publishing, 2021 – 140p.
6. Mar, G.R., Grim, P., Denis, P.S. *The Philosophical Computer: Exploratory Essays in Philosophical Computer Modeling*. London: MIT Press, 1998 – 333p.
7. Mayo-Wilson, C., Zollman, K.J.S. The computational philosophy: simulation as a core philosophical method // *Synthese*. 2021. Vol. 199. P. 3647–3673.
8. Rugai, N. *Computational Epistemology: From Reality To Wisdom*. Brooklyn: Lulu.com Press, 2012. – 304 p.

9. Thagard, P. *Computational Philosophy of Science*. Cambridge, Massachusetts: Bradford Books, 1993– 256 p.

10. Vámos, T. *Knowledge and Computing: Computer Epistemology and Constructive Skepticism*. – Budapest: Central European University Press, 2010. - 210 p.

11. Aleksandr Lojko. *Biases of natural and artificial intelligence: Social costs and risks*. – Our Knowledge Publishing, 2024 – 96p.

12. Loiko, A.I. Digital colonialism and national movements for digital independence // *Bulletin of Udmurt University. Sociology. Political Science. International Relations*. 2026. Vol. 10, iss. 1. P. 52–56. <https://doi.org/10.35634/2587-9030-2026-10-1-52-56>