English original: Roman Lukyanenko. Next Data Paradigm: Using AI to Manage All Human Data. Foundations, Architecture, and Challenges in Using a Universal AI Data Manager. 2025. hal-05028057, https://hal.science/hal-05028057v1

Искусственный интеллект для управления всеми человеческими данными

Роман Лукьяненко
Школа коммерции Макинтайра
Университет Вирджинии
romanl@virginia.edu

Абстрактный
В эпоху информационных технологий управление данными — основа информационных технологий — остается трудоемким, неэффективным, в значительной степени недоступным, далеким от своего потенциала. Средства для значительного скачка вперед в управлении данными уже здесь. Стремительное развитие искусственного интеллекта представляет собой возможность смены парадигмы в цифровом хранении и управлении данными. В этой статье рассматривается, как системы агентного (искусственного интеллекта) ИИ могут революционизировать способы хранения, организации и извлечения данных организациями и людьми. Мы предлагаем ИИ для управления всеми потребностями людей в хранении и извлечении данных. Используя передовые возможности машинного обучения и автономного принятия решений, управление данными на основе ИИ обещает превратить управление данными из неэффективного, требующего много времени процесса в интеллектуальную персонализированную услугу, доступную каждому.

Ключевые слова
Управление данными, искусственный интеллект, агентный ИИ, генеративный ИИ, этика ИИ, большие языковые модели, большие модели рассуждений, большие модели действий, моделирование данных, сбор данных, управление данными, инфраструктура данных, аналитика данных, инклюзивное управление данными, экологичное управление данными

1. Цифровой неолит, поздние данные каменного века

От беспилотных автомобилей до реалистичных фейков, созданных искусственным интеллектом, до гипераддиктивных видеороликов на YouTube, TikTok и Netflix, до полезных инструментов, таких как MS Word и ChatGPT, прогресс информационных технологий ошеломляет. Темпы развития событий завораживают: «год в машинном обучении [основа современного искусственного интеллекта] равен столетию в любой другой области» [151] .

Удивляясь цифровому миру, трудно поверить, что наш мир, вероятно, является каменным веком данных. Вы правильно прочитали: каменный век. Как это может не быть вершиной человеческого развития, можете вы возразить? И все же, со всеми достижениями в области ИИ и других технологий, мы, вероятно, находимся только на заре ИТ-революции. Причина в том, что управление данными, сама основа информационных технологий, остается трудоемкой, неэффективной, в значительной степени недоступной и далеко не реализующей свой потенциал.

Управление данными есть везде. Организации, сотрудники и люди, занимающиеся своей повседневной жизнью, постоянно управляют цифровыми данными. И почти каждый раз это борьба. Сохранить фотографию на смартфоне, зарегистрироваться на концерт, сделать публикацию в социальных сетях, создать резервную копию рабочих файлов в облаке — все это несложно. Вспомнить детали в электронном письме, полученном неделю назад, выяснить последнюю версию общего документа, найти все соответствующие данные клиентов, защитить личные записи от злоумышленников — все это не так-то просто. Как будто по какой-то жестокой иронии мы легко создаем цифровые данные, только чтобы потом с трудом ими управлять.

Хотя вычислительные устройства, облако, искусственный интеллект, платформы и приложения достигли значительного прогресса, большинство задач по управлению данными в значительной степени зависят от человеческого вмешательства и крайне неэффективны. Управление данными обычно является синонимом усилий и часто разочарования. В некоторых проектах «обнаружение данных, прием данных, очистка данных и проектирование конвейера данных» могут занять «месяцы» [64] . Было подсчитано, что 80% времени в аналитических проектах тратится на управление данными [60] .

Плохое управление данными приводит к огромным финансовым и репутационным потерям. Даже ведущие компании в сфере информационных технологий испытывают трудности с управлением данными, о чем свидетельствуют постоянные громкие провалы. Рассмотрим, например, провал IBM Watson - MD Anderson Cancer Center, который вместо лечения рака привел к многомиллионным убыткам [144, 161] . Из-за частых провалов ценные возможности обычно остаются неиспользованными.

Только в США предполагаемая стоимость плохого качества данных превышает 3 триллиона долларов [53, 140] . Говорят, что организации тратят до 30% своих доходов на решение проблем качества данных [53] . Эти шокирующие цифры являются ярким свидетельством того, насколько важными стали данные и насколько нетривиальной стала задача управления ими.

Хорошей новостью является то, что мы находимся в цифровом неолите, позднем каменном веке данных. Это означает, что перемены близки. Искусственный интеллект, использующий такие достижения, как генеративный ИИ [8, 162] и агентный ИИ [14, 54], является естественным ответом на натиск данных. Это рецепт развития в другую парадигму управления данными, чтобы возвестить о новой эре данных.

В этой статье предлагается ключевая особенность Next Data Age: управление данными на основе ИИ. Идея проста: для всех потребностей в данных (создание, сбор, хранение, извлечение данных) есть персональный исполнитель ИИ, который будет ими управлять. Исполнитель данных ИИ (AI Data Executive, AIDE) динамически подстраивается под предпочтения, навыки и возможности пользователя и обрабатывает все файлы, документы и информацию, которые предоставляет пользователь. Он абстрагирует процесс обработки и хранения, который в высокой степени оптимизирован для обеспечения интеллектуального извлечения и экологической эффективности. Когда пользователь хочет увидеть или обменяться данными, AIDE извлекает и предоставляет их. Однако это простое видение таит в себе много сложностей, включая множество технических, социальных, этических проблем и

проблем безопасности, которые необходимо решить. В статье излагаются общие принципы AIDE и предлагаются открытые вопросы, на которые необходимо ответить, чтобы сделать управление данными ИИ основным элементом Next Data Age.

2. Современное управление данными

2.1. Проблемы и недуги современного управления данными

Управление данными — трудоемкое, неэффективное и далеко не реализующее свой потенциал. Проблема усугубляется со временем, поскольку объемы данных растут, а зависимость от цифровых данных увеличивается. Вот некоторые наблюдения о состоянии управления данными сегодня (обобщенные на рисунке 1).

Начнем с необычной проблемы управления данными, которая, к сожалению, не получает должного внимания. Начнем с вопроса включения.

Включение в управление данными гарантирует, что любой заинтересованный в сборе, хранении и использовании цифровых данных может делать это с легкостью и выгодой. Современное управление данными не является инклюзивным. Точка. Рассмотрим лишь несколько примеров недоступности управления данными. Большая часть управления данными (например, настройка интерфейсов сбора данных, создание моделей данных, написание схем баз данных, оптимизация запросов) в значительной степени опирается на визуальную модальность человека. Это создает значительный барьер для людей с нарушениями зрения. Кроме того, управление данными, особенно в масштабе, требует значительных технических навыков, таких как знание специализированного программного обеспечения (например, Hadoop, NoSQL, озера данных) и языков (например, моделирование данных, Python, SQL). Некоторые группы постоянно маргинализируются и продолжают плохо представляться в управлении данными (например, женщины). Определенные виды деятельности по управлению данными [49, 53, 110] , такие как моделирование, сбор и получение данных, организация и курирование данных, были отмечены как плохо доступные для различных пользователей [12, 41, 56, 69, 72, 82–84, 100, 103, 107, 111, 118, 122, 145, 150, 152, 187] .

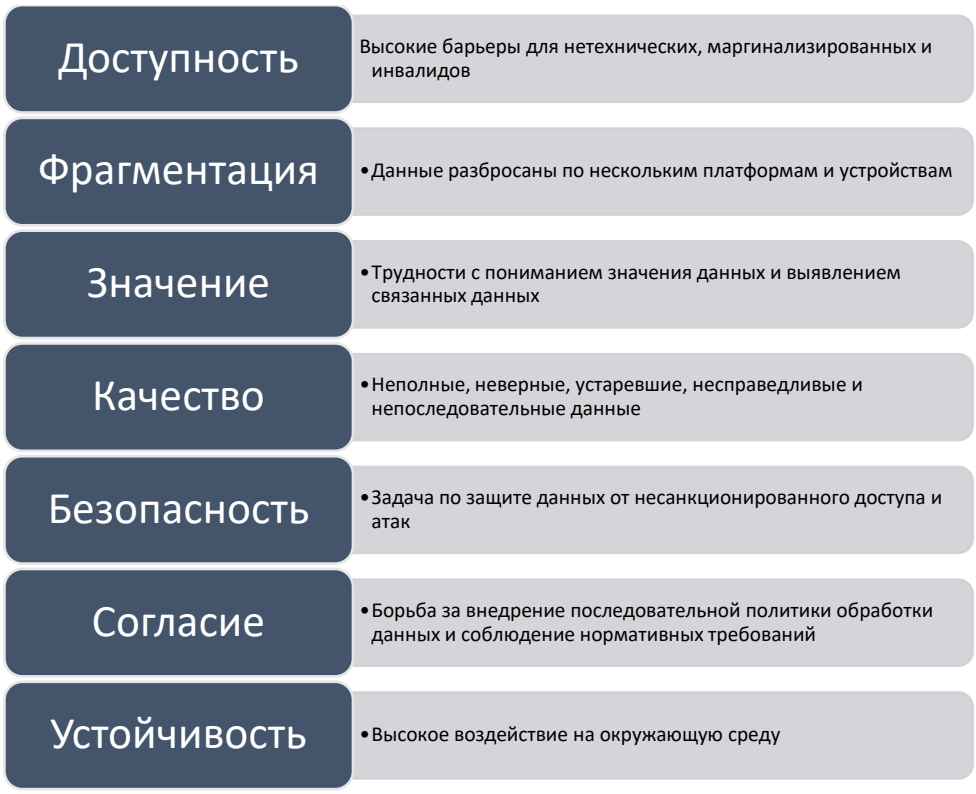

Рисунок 1. Проблемы управления данными в позднем каменном веке

Инклюзивное управление данными — это управление данными, которое учитывает различные уровни навыков и доступа во всех видах деятельности по управлению данными. Это включает все аспекты управления данными, такие как моделирование, приобретение, управление, инфраструктура и поддержка потребления (так называемая MAGIC управления данными). В настоящее время основное внимание в доступности уделяется вопросам дизайна интерфейса (аспект приобретения данных MAGIC). Напротив, вопросы, связанные с моделированием, управлением, инфраструктурой и поддержкой потребления, обычно игнорируются. Снижая барьеры для доступа, инклюзивное управление данными должно обеспечивать более справедливое и эффективное использование цифровой информации, чтобы каждый мог воспользоваться богатством возможностей цифрового мира.

Фрагментация данных создает множество проблем для людей и организаций. Файлы и информационные записи разбросаны по множеству платформ и устройств — персональным компьютерам, облачным дискам, мобильным телефонам и организационным системам хранения данных. Более того, данные — это не только отдельные документы, изображения или видеофайлы. Это также информация, хранящаяся внутри цифровых систем, таких как платформы, приложения, программные пакеты, разбросанные по цифровой вселенной. Следовательно, данные о клиентах в организации могут быть распределены по системе управления взаимоотношениями с клиентами, платформе маркетинга по электронной почте, электронным таблицам продаж и журналам обслуживания клиентов.

Такая фрагментация увеличивает риски и порождает неэффективность. Финансовые отчеты могут опираться на данные из бухгалтерского программного обеспечения, платформ управления цепочками поставок и прогнозов продаж — каждый из которых хранится в разных форматах и обновляется по отдельным графикам. Одно несоответствие между этими источниками может исказить весь финансовый прогноз, но их согласование требует ручных усилий и межведомственной координации. В

результате организация может упустить критически важный контекст, например недавнюю жалобу клиента или ожидаемое продление контракта, что приведет к плохому обслуживанию и упущенным возможностям для бизнеса (или к чему-то похуже).

На личном уровне люди сталкиваются с похожими проблемами при попытке управлять своими личными данными. Человек, планирующий поездку, может иметь данные о рейсах в своей электронной почте, бронирование отелей, сохраненное в виде PDF-файлов на облачном диске, и заметки о достопримечательностях в мобильном приложении. Если человек хочет создать единый маршрут или поделиться планом с другом, будет мучением вручную собирать эти данные с разных платформ и в разных форматах. Хуже того, если какие-либо данные устарели или сохранены в формате, который больше не совместим с их текущим устройством или программным обеспечением, им, возможно, придется вручную конвертировать их, если это вообще возможно.

На работе или дома людям нужно помнить, где они хранят определенные файлы, и вручную обновлять и переносить документы, изображения, видео между системами. Синхронизация между платформами часто непоследовательна, заставляя людей отслеживать разные версии файлов и вручную разрешать конфликты. Извлечение данных зависит от точного наименования или структуры папок, а функции поиска часто ограничены неполными метаданными или плохой индексацией, что затрудняет быстрый поиск нужной информации.

Отсутствие контекстной осведомленности о данных службами управления данными является основной причиной усилий и неудач. Системы данных с трудом понимают глубокий смысл данных и идентифицируют семантические связи. Текущие инструменты управления данными могут хранить и извлекать файлы, но с трудом распознают связи между различными наборами данных или форматами. Например, презентация, электронная таблица с финансовыми данными и электронное письмо, в котором обсуждается один и тот же проект, могут храниться в совершенно разных местах без какой-либо логической связи между ними. Пользователи должны вручную объединять и перекрестно ссылаться на эти файлы, что увеличивает риск упущения важной информации или принятия решений на основе неполных данных. Кроме того, категоризация и маркировка данных по-прежнему в значительной степени зависят от ввода данных пользователем, который непоследователен и подвержен ошибкам [6, 74, 101, 154] .

Низкое качество данных — в результате неполных, устаревших или непоследовательных данных — может подорвать принятие решений и действия, предпринимаемые с данными. Такие проблемы, как дублирование записей, неверная информация о клиентах и неправильно классифицированные данные, увеличивают риски и снижают операционную эффективность [36, 93, 141] . Даже ведущие компании в сфере информационных технологий сталкиваются с проблемами качества данных. Таким образом, после многомиллионных инвестиций и большой шумихи онкологический центр MD Anderson Cancer Center отказался от сотрудничества с системой ИИ Watson, предоставленной IBM [144] . И это несмотря на амбициозную цель IBM использовать ИИ для лечения рака. На крах проекта MD Anderson-Watson существенно повлияли проблемы качества данных. Проект боролся с несоответствиями и ошибками в данных, что препятствовало эффективности системы Watson в предоставлении точных и безопасных рекомендаций по лечению рака [161] .

Качество данных — это не только техническая проблема обеспечения точности, полноты или своевременности данных — общие соображения по качеству данных [16, 36, 101, 104, 109, 124, 164, 176, 178, 184] . Данные — это социальный артефакт [1, 2, 7, 132], и качество также подразумевает справедливость — были ли данные собраны

этично, законно и в соответствии с соответствующими культурными нормами. Современное понимание справедливости данных разбросано по конкретным областям внимания, таким как вопросы информированного согласия в научных исследованиях и опросах потребителей [33, 50, 92] или допустимость юридических доказательств [67, 86, 94, 123] . Справедливость недостаточно хорошо включена в общие структуры управления данными и, следовательно, может быть легко упущена из виду.

Безопасность данных еще больше усложняет управление данными. Обеспечение безопасности данных остается чрезвычайно сложной задачей, поскольку объемы и разнообразие данных растут. Число атак и утечек данных растет, что приводит к значительному ущербу. Прогнозируется, что глобальные затраты на киберпреступность составят 10,5 триллионов долларов [120] , ежегодно увеличиваясь на 15 процентов [25] . Например, утечка данных, с которой столкнулась Equifax в 2017 году, была вызвана неадекватными мерами безопасности данных, которые позволили хакерам получить доступ к конфиденциальной информации примерно 147 миллионов человек. Количество кибератак с использованием украденных учетных данных увеличилось на 71% по сравнению с прошлым годом [25] .

Распространение устройств и данных в различных системах усугубляет риски безопасности, поскольку пользователи прибегают к небезопасным обходным путям, таким как отправка файлов по электронной почте себе или использование несанкционированных платформ обмена файлами. Организациям и отдельным лицам приходится ориентироваться на сложные настройки разрешений для безопасного обмена файлами между командами или личными устройствами, чтобы предотвратить несанкционированный доступ и атаки. Процессы резервного копирования и восстановления также разрознены — большинство систем полагаются на запланированное резервное копирование, а не на сохранение критически важных данных в реальном времени с учетом контекста.

С взрывным ростом данных и прогрессом в области ИТ безопасность становится все сложнее контролировать даже опытным ИТ-специалистам. Например, появление квантовых вычислений грозит подорвать криптографические основы современной безопасной связи. Исследователи, компании и регулирующие органы по всему миру пытаются решить надвигающиеся проблемы квантовой безопасности [20, 47, 81, 81, 137, 142] .

Соблюдение существующих правил и положений является особенно сложной задачей для организаций (и, в меньшей степени, для отдельных лиц). По мере того, как растет важность данных в обществе, растут и правила и положения, которым организации должны следовать, чтобы не столкнуться с санкциями. Многие организации испытывают трудности с внедрением четких политик в отношении владения данными, прав использования и хранения, что приводит к несанкционированному доступу к данным, непоследовательному качеству данных и несоблюдению таких норм, как GDPR (Общий регламент по защите данных) и CCPA (Закон Калифорнии о защите конфиденциальности потребителей). Неспособность обеспечить надлежащее управление может привести к юридическим санкциям, репутационному ущербу и потере доверия клиентов.

Воздействие управления данными на окружающую среду ошеломляет и продолжает расти. Это неудивительно. Трудно эффективно управлять фрагментированными данными. В то время как для большинства людей ИТ невидимы (некоторые даже называют их виртуальными, как будто их не существует) [113] , все ИТ являются физическими, и эта физическая вселенная данных наносит очень большой урон окружающей среде [75, 98, 133] . Большая часть этого урона приходится на управление данными [106] . Таким образом, только на облачные хранилища приходится более 1% мирового потребления электроэнергии [73] . Если бы «цифровой

мир был страной, он был бы третьим по величине потребителем энергии после Китая и США» [135] .

На рисунке 2 современные проблемы управления данными визуализированы в виде облака слов.

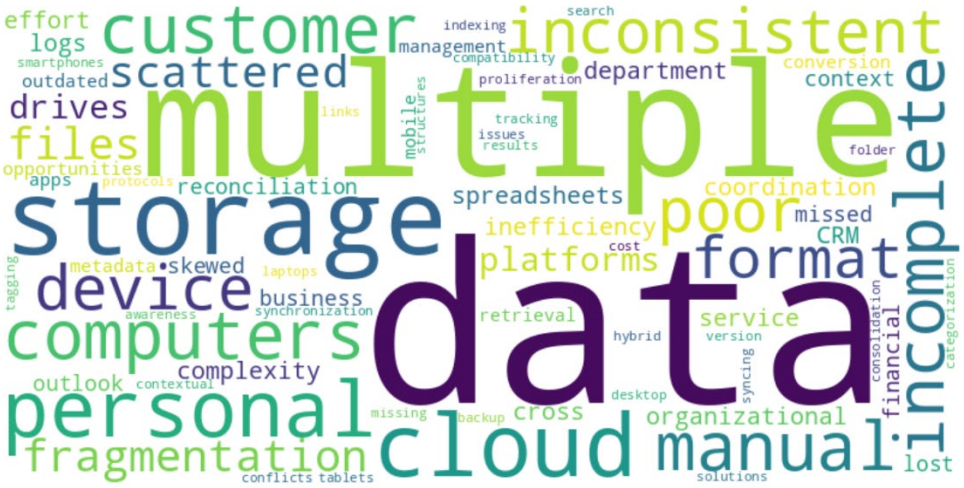

Рисунок 2. Проблемы управления облаком данных

2.2. Причины проблем управления данными

Сложность, которая превышает человеческие возможности, является фундаментальной причиной борьбы с данными, с которой мы сталкиваемся сегодня. Пока создавать новые данные легко, и это полезно и приятно, люди будут продолжать это делать, что приведет к увеличению объемов данных. Чем больше цифровых данных формируют различные грани человеческого существования, тем больше производится разнообразных видов данных, чтобы соответствовать сложности окружающего нас мира. Чем больше объем и разнообразие, тем сложнее становится организация, курирование и последующее извлечение данных. Yahtzee! Мы участвуем в гонке со сложностью.

Люди по своей природе склонны документировать, делиться и хранить информацию как средство общения, решения проблем и самовыражения [52, 115, 116] . Распространение смартфонов, платформ социальных сетей и облачных приложений сделало создание данных практически беспрецедентным, ускорив рост объемов данных. Чем более доступными и интегрированными становятся эти платформы, тем больше людей и организаций вносят вклад в экосистему данных, усиливая цикл обратной связи непрерывного создания данных.

Чем больше цифровых данных внедряется в различные аспекты человеческого существования, тем более разнообразными и специализированными становятся типы данных. Например, бизнес-данные теперь включают в себя не только структурированную информацию, такую как финансовые записи, но и неструктурированные данные, такие как отзывы клиентов, взаимодействия в социальных сетях и мультимедийный контент [3, 37, 38, 42, 46, 85] . Аналогичным образом, персональные данные охватывают не только текстовую информацию, но и показатели здоровья с носимых устройств, данные о местоположении с GPS-отслеживания и изображения с социальных платформ. Разнородность этих данных создает значительные проблемы для интеграции, поскольку разные форматы и источники требуют разных методов обработки и обращения. Без интеллектуальных систем, способных обрабатывать такое разнообразие, задача организации и извлечения данных становится все более трудоемкой и неэффективной.

Проблема еще больше усугубляется тем фактом, что создание данных часто опережает инфраструктуру управления данными. Хотя емкость хранения и возможности обработки данных улучшились, они не успевают за экспоненциальным ростом объема и разнообразия данных. В результате организации и отдельные лица часто сталкиваются с хранилищами данных, непоследовательными файловыми структурами и несовместимыми форматами. Рост облачных вычислений и удаленной работы усугубил эту проблему, поскольку данные теперь распределены по нескольким платформам и географическим местоположениям. Эта децентрализация создает пробелы в видимости и контроле данных, что затрудняет установление единого источника истины. Чем сложнее становится среда данных, тем больше когнитивная и операционная нагрузка на пользователей по навигации и согласованию разрозненных источников данных.

Когнитивная нагрузка, связанная с организацией и поиском данных, увеличивается с разнообразием источников и форматов данных. Рабочая память человека ограничена в своей способности обрабатывать и категоризировать информацию [9, 119] . При столкновении с чрезмерными или плохо организованными данными люди испытывают повышенное умственное напряжение, что приводит к более медленному принятию решений и более высокому уровню ошибок [80, 91, 131] . В организационных условиях это проявляется в снижении производительности, недопонимании и стрессе, особенно для пожилых сотрудников [160, 166, 167] или сотрудников с ограниченными возможностями [165] . Даже на индивидуальном уровне усилия, необходимые для консолидации персональных данных на разных устройствах и платформах, например, объединение данных о состоянии здоровья с умных часов с записями о питании из мобильного приложения, создают барьеры для эффективного использования данных и принятия решений.

Подводя итог, можно сказать, что пока создание данных остается простым и полезным, объем и разнообразие данных будут продолжать расти, усиливая сложность экосистемы данных. Поэтому эффективные стратегии управления данными должны учитывать как человеческую тенденцию к созданию данных, так и человеческие ограничения в управлении этой сложностью. Решение современных задач управления данными требует большего, чем просто расширение хранилища или повышение скорости обработки; это требует разработки интеллектуальных систем управления данными, способных к контекстному пониманию и адаптивной организации.

3. ИИ — лучший менеджер данных

3.1. Фонды

В этой статье выдвигается критическое предложение для Next Data Age: управление данными на основе ИИ как фундаментальная парадигма данных. В основе этого видения лежит концепция AI Data Executive (AIDE) — персональной адаптивной системы, предназначенной для автономного управления всеми аспектами взаимодействия с данными, включая создание, сбор, хранение и извлечение. Динамически подстраиваясь под предпочтения, навыки и возможности пользователя, AIDE берет на себя полную ответственность за организацию и обработку файлов, документов и информации, тем самым абстрагируя сложности управления данными в бесшовный, оптимизированный процесс. Эта интеллектуальная система не только повышает эффективность извлечения данных, но и отдает приоритет экологической устойчивости с помощью механизмов хранения с учетом ресурсов.

Несмотря на привлекательность, реализация AIDE влечет за собой значительные технические, социальные, этические проблемы и проблемы безопасности, которые

необходимо тщательно решать. Прежде чем сформулировать архитектурные принципы AIDE, мы четко излагаем ее основные ценности. Это вдохновлено гибкой разработкой и Agile Manifesto [17] . Может быть несколько способов реализации AIDE, и мы не предлагаем, чтобы наш был окончательным или лучшим. [1]Однако, излагая основополагающие ценности системы, мы вдохновляем будущие разработки, которые могут найти лучшие способы реализации этих ценностей в конкретных технологических реализациях. Мы предлагаем искусственный интеллект для управления всеми потребностями людей в хранении и извлечении данных. Это необходимо для обработки чрезмерной и растущей сложности управления данными. Полученный в результате AI Data Executive (AIDE) представляет собой конкретную систему, разработанную для достижения этих целей. AIDE или любая подобная система должна придерживаться следующих основополагающих ценностей:

Ценность 1. **Человек прежде всего, человек — цель, а не средство.**

Ценность «Человек прежде всего, человек как цель, а не средство» отражает философскую и этическую позицию, которая ставит во главу угла человеческое благополучие, достоинство и автономию в любой системе, технологии или общественной структуре. Она предполагает, что:

1. Человек всегда в приоритете. Человеческие потребности, права и ценности должны быть главными соображениями при любой реализации ИИ. Вместо того, чтобы отдавать приоритет эффективности, прибыли или автоматизации ради самих себя, системы ИИ, такие как AIDE, должны быть разработаны для обслуживания и расширения прав и возможностей людей.
2. «Человек как цель, а не средство. Пользователи-люди никогда не должны рассматриваться просто как инструменты или ресурсы для достижения внешних целей (например, корпоративной прибыли, технологического прогресса или политической власти). Вместо этого все системы и достижения должны в конечном итоге служить процветанию человека, а не низводить людей до инструментов для какой-то другой цели.

Эта ценность берет свое начало в кантовской этике [180] , которая утверждает, что к людям всегда следует относиться как к цели самой по себе, а не просто как к средству достижения чужой цели.

Ценность 2. **Безопасность, надежность, устойчивость, прозрачность по сравнению с чистой производительностью.**

Ценность «Безопасность, надежность, прозрачность вместо чистой производительности» предполагает, что при проектировании и развертывании ИИ определенные основополагающие ценности должны иметь приоритет над скоростью, эффективностью или вычислительной мощностью.

---

[1]В целом, существует множество способов реализовать любую идею в технологическом решении, проблема, известная как эквифинальность (заданное конечное состояние или результат могут быть достигнуты многими различными путями) [21, 65, 70, 112, 158] . Цель состоит в том, чтобы найти тот, который наилучшим образом соответствует контекстным требованиям, при этом максимально соблюдая ценности AIDE.

1. Безопасность прежде всего. Системы должны быть спроектированы так, чтобы предотвращать вред, как физический, так и цифровой. ИИ и автоматизированные системы не должны представлять риска для отдельных лиц, общества или окружающей среды. Безопасность включает в себя устойчивость к сбоям, враждебным атакам и непреднамеренным последствиям.
2. Надежность. Система должна функционировать последовательно и предсказуемо в различных условиях. Она не должна вести себя хаотично, совершать критические ошибки или требовать чрезмерного вмешательства человека для исправления непреднамеренных результатов.
3. Устойчивость. При реализации следует отдавать приоритет влиянию на окружающую среду в целом, включая животных и природные ресурсы планеты.
4. Прозрачность. Внутренние механизмы, процессы принятия решений и ограничения системы должны быть понятны пользователям и заинтересованным сторонам. Модели искусственного интеллекта «черного ящика» или непрозрачные механизмы принятия решений могут привести к недоверию, предвзятости и непреднамеренному вреду. Прозрачность обеспечивает подотчетность и способствует ответственному использованию. Естественно, мы признаем невозможность полной прозрачности, особенно при работе с высокопроизводительными моделями искусственного интеллекта (например, нейронными сетями глубокого обучения) [4, 4, 161] .

Это значение согласуется с основами дизайна, как изложено в работе Ларсена и др. [89] о валидности дизайна с научной точки зрения. Любой надежный дизайн в соответствии с валидностью дизайна с научной точки зрения должен не только демонстрировать превосходную производительность (валидность критерия), но и проектировщики должны уверенно знать, как работает система (каузальная валидность) и где ее целесообразно использовать (контекстная валидность).

Значение 3. **Владельцы данных должны иметь полный контроль над своими данными.**

Ценность «Владельцы данных должны иметь полный контроль над своими данными» пропагандирует суверенитет данных, который утверждает, что лица или организации, владеющие данными, имеют право управлять, получать доступ и определять, как их данные используются, передаются и распространяются. Эта ценность выступает за личную автономию над цифровыми данными, гарантируя, что владельцы сохраняют полную власть над решениями относительно своих данных, а не позволяют третьим лицам или внешним организациям диктовать их использование без согласия. Это включает возможность предоставлять или отзывать доступ, изменять данные или удалять их, сохраняя при этом прозрачность в отношении того, как обрабатываются данные.

Рисунок 3. Основополагающие ценности AIDE

Три основополагающие ценности AIDE (Рисунок 3) способствуют подходу, ориентированному на человека, к технологиям и управлению данными. Вместе эти принципы выступают за этичный, ответственный и ориентированный на человека технологический прогресс. Теперь мы покажем, как эти ценности реализуются.

3.2. Архитектурные принципы

Общая архитектура AI Data Executive (AIDE) представлена на рисунке 4. Она отображает базовую настройку, которая может быть специализирована для конкретных нужд, отраслей или в соответствии с конкретными нормативными требованиями. Цель любой архитектуры — как можно лучше реализовать три основополагающих ценности, лежащие в основе AIDE. Общие советы о том, как внедрять абстрактные идеи в конкретные технологические реализации (или устанавливать валидность инстанцирования), см. в [10, 43, 44, 89, 99, 102, 105, 112] .

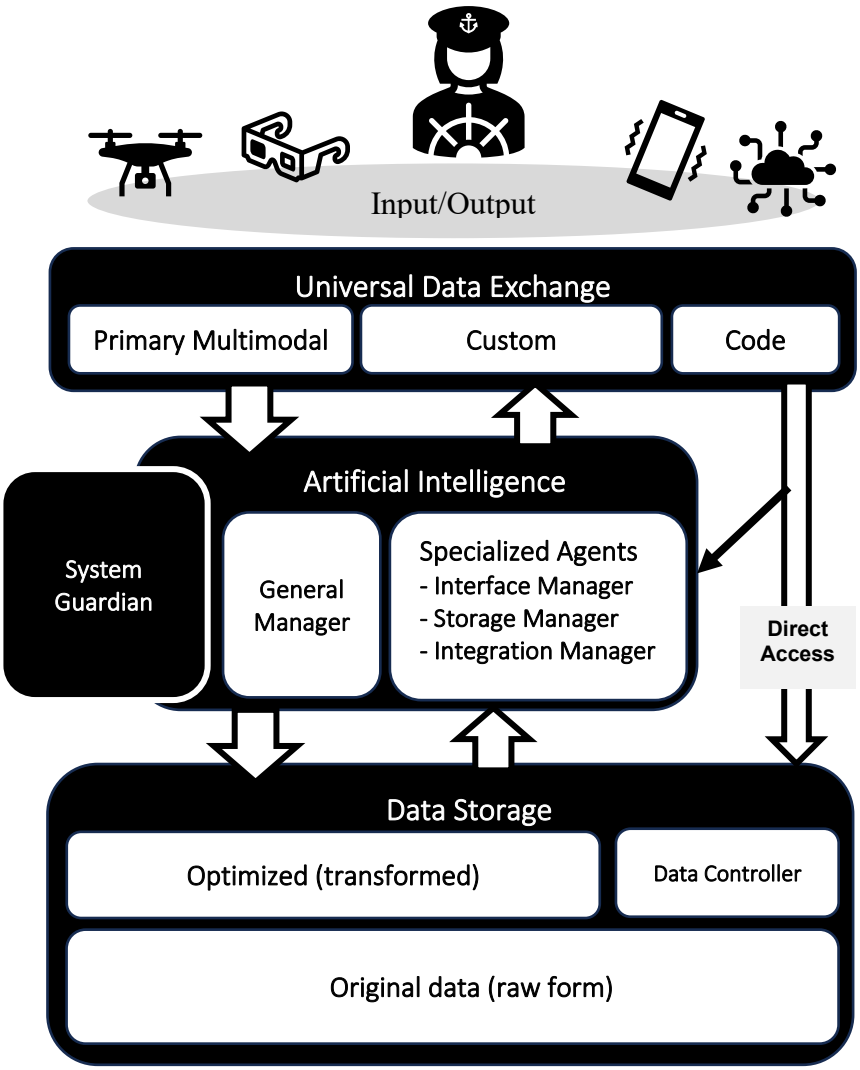

Рисунок 4. Архитектура менеджера данных ИИ

3.2.1. Владелец системы

Человеческая фигура наверху — это Владелец **системы** , человек, группа людей или их назначенные доверенные лица, которые владеют данными. Напомним, что основополагающая ценность заключается в том, чтобы Владелец системы всегда полностью контролировал свои данные, всегда был полностью информирован о любом принятом решении и чтобы ни одно решение не могло быть принято без осознанного согласия Владельца системы. Остальная часть архитектуры реализует эти идеалы.

AI Data Executive — это многоуровневая модульная система. Это обеспечивает ее способность развиваться и быть гибкой в соответствии с меняющимися потребностями пользователей, технологическими разработками и нормативными требованиями [51, 87, 111, 125, 153, 188] . Данные передаются с поддержкой ИИ от источника — создателя данных, будь то человек или контролируемый человеком процесс — к хранилищу. Данные также передаются обратно потребителю данных — владельцу или устройствам, уполномоченным владельцем системы получать определенную информацию.

Владелец системы может свободно использовать широкий спектр устройств, таких как смартфоны, ноутбуки, дроны, 3D-принтеры и тактильные интерфейсы, как для

сбора, так и для отображения информации. Все это управляется и хранится централизованно, с помощью искусственного интеллекта.

В то время как владелец системы имеет возможность выбирать, какое устройство использовать, устройства накладывают физические ограничения на тип информации, которая может быть собрана и отображена. С другой стороны, свойства устройств делают возможным определенный тип сбора и представления данных. Например, смартфон может собирать данные о местоположении, дрон может делать аэрофотоснимки, а 3D-принтер может создавать физические модели на основе обработанных данных.

Сами устройства могут быть или не быть частью AIDE. Возможны различные варианты. Следовательно, можно разработать некоторые устройства, которые предоставят дополнительные возможности для использования преимуществ AIDE. Например, дрон может быть специально разработан для дополнения AIDE, что позволит ему отправлять изображения ландшафта в реальном времени непосредственно в систему для анализа и передачи на другие устройства (например, 3D-принтер). Таким образом, владелец системы может организовать многоступенчатый обмен информацией для автоматизации некоторых сложных задач (например, создание 3D-изображения ландшафта для строительных целей). В то же время общие вычислительные устройства, такие как ноутбуки и смартфоны, могут просто установить AIDE как программный пакет или приложение, тем самым позволяя интегрировать эти устройства в экосистему AIDE.

В устройстве, работающем под управлением AIDE, собранные данные обрабатываются и стандартизируются с помощью Universal Data Exchange. Таким образом, данные остаются доступными и совместимыми на нескольких устройствах. Это позволяет AIDE управлять данными согласованно, даже если устройства различаются по возможностям и конструкции.

3.2.2. Универсальный обмен данными

Вся собранная информация передается через **универсальный обмен данными** (UDE), который стандартизирует и переводит данные в унифицированный формат хранения. UDE действует как центральный узел, облегчая интеграцию данных и повышая общую эффективность и адаптивность системы. Выходные возможности UDE преобразуют данные в применимые форматы, соответствующие функциональности принимающего устройства. Например, данные, собранные со смартфона о качестве воздуха, можно визуализировать на панели управления ноутбука или преобразовать в тактильную обратную связь на тактильном интерфейсе для оповещения пользователей с нарушениями зрения об опасных условиях.

Универсальный **обмен данными** состоит из трех компонентов: основной мультимодальный, пользовательский и кодовый. При использовании различных устройств владелец системы может использовать определенные функции этих трех компонентов при условии, что устройства допускают интеграцию с этими компонентами. Не все устройства смогут поддерживать каждый компонент. Например, некоторые устройства могут не позволять пользователю напрямую писать код или манипулировать пользовательскими объектами графического интерфейса. В этом контексте роль основного мультимодального компонента, который является компонентом на базе ИИ, заключается в автоматическом обнаружении физических возможностей каждого устройства и соответствующей настройке параметров сбора и представления данных . Напротив, компоненты кода и пользовательского компонента не основаны на ИИ и могут потребовать от владельца системы вручную согласовывать возможности конкретного устройства для включения этих компонентов.

Primary Multimodal — это основной интерфейс для сбора и представления данных. Он управляется ИИ и стремится собирать и предоставлять данные в любом возможном режиме. Целью является поддержка известных сенсорных систем человека [159] :

- **Зрение** (визуальная система): Графический интерфейс использования, стандартный и наиболее используемый тип интерфейса сегодня. Установлены принципы взаимодействия пользователя с графическими интерфейсами [155, 168] . Достижения в области ИИ, включая обработку естественного языка и компьютерное зрение, обеспечивают большую адаптивность и настраиваемость графического интерфейса [156] .
- **Слух** (слуховая система): голосовой и звуковой интерфейс. Более новый, но теперь становится широко доступным благодаря прогрессу в распознавании речи, преобразовании текста в речь, обработке естественного языка [156] .
- **Вкус** (вкусовая система): Вкусовой интерфейс в настоящее время является футуристической идеей предоставления реального опыта вкуса. Моделирование вкуса уже разрабатывается [126, 174] .
- **Обоняние** (обонятельная система): Обонятельный интерфейс позволяет обнаруживать запахи и производить запахи. Такой интерфейс уже рассматривается [126] .
- **Осязание** (тактильная система): позволяет пользователям чувствовать такие ощущения, как давление, температура и боль через рецепторы кожи. Эти интерфейсы появляются, включая генерацию 3D-поверхностей, системы Брайля и реалистичные видеоигры [18] .

Основной мультимодальный интерфейс AIDE позволяет пользователям взаимодействовать с системой ИИ с помощью различных модальностей в зависимости от контекста, предпочтений и способностей пользователя. Интерфейс динамически подстраивается под стиль ввода пользователя и поставленную задачу. Например, пользователь может начать взаимодействовать с системой через графический интерфейс, такой как панель инструментов, отображающая визуализацию данных. Если пользователю нужны более глубокие знания, он может переключиться на естественный язык, задавая уточняющие вопросы на простом языке, например: «Почему продажи упали в прошлом квартале?» Этот переход между графическим и естественным языковым вводом позволяет интерфейсу адаптироваться к различным уровням технической компетентности и предпочтительным способам обработки данных.

Необязательным компонентом AIDE является **интерфейс мозг-компьютер** (BCI), система, которая устанавливает прямой путь связи между мозгом и внешним устройством, таким как компьютер или протез конечности. BCI работают, обнаруживая и интерпретируя нейронные сигналы, как правило, электрическую активность мозга, и преобразуя их в команды, которые управляют внешним устройством или программным обеспечением. Целью BCI является предоставление людям возможности взаимодействовать с машинами или средами, не полагаясь на традиционные мышечные пути. Понятно, что BCI является спорной технологией и подходит не всем. Однако BCI играет важную роль в поддержке людей с некоторыми ограниченными возможностями и тех, кто проходит реабилитацию, которую можно облегчить с помощью BCI. Поэтому AIDE поддерживает BCI.

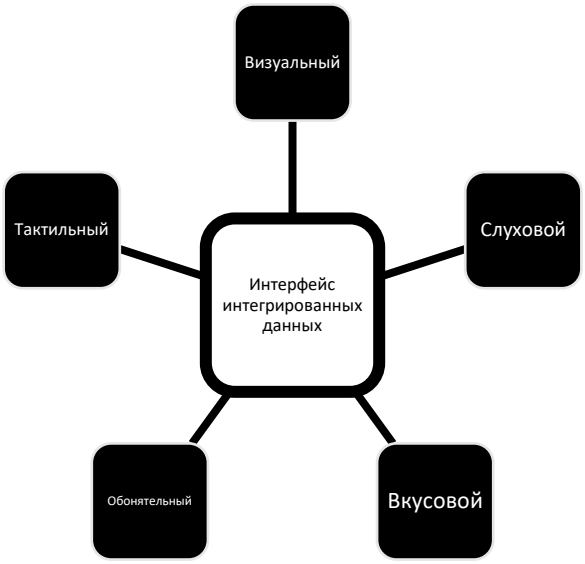

Рисунок 5. Интегрированный сенсорный интерфейс UDE

Оптимальная реализация AIDE заключается в обеспечении интеграции всех сенсорных систем (см. Рисунок 5). Таким образом, интерфейс, управляемый ИИ, может адаптироваться на основе факторов окружающей среды и ситуации. Например, в производственных условиях техник может использовать тактильную обратную связь для регулировки настроек оборудования напрямую через тактильный интерфейс, а также получать диагностические отчеты в реальном времени через графический или естественный языковой дисплей. Только интегрированный интерфейс позволит людям с разными навыками, способностями и потребностями эффективно управлять данными.

Важным компонентом AIDE является **Custom Interface** . В отличие от Primary Multimodal, Custom Interface не управляется адаптивностью AI, а скорее является чистым листом, позволяя пользователю разрабатывать собственные интерфейсы, настраивая и фиксируя модальности и представления по своему выбору. Этот интерфейс лишен адаптивности Primary Multimodal, но обеспечивает возможность System Owner иметь полный контроль над всеми компонентами интерфейса. AI по-прежнему будет отвечать за обработку данных, сгенерированных Custom Interface. Вывод данных контролируется пользователем и может быть представлен либо через Custom Interface, либо через Primary Multimodal, в зависимости от типа данных и правил, установленных System Owner.

Оба интерфейса Primary и Custom способны выполнять программный код, такой как Python или SQL. Однако Primary Multimodal interface управляется ИИ, и, следовательно, интерпретация и выполнение кода опосредованы ИИ. Для обеспечения прямого контроля над данными с помощью кода предоставляется **Code Interface** . Интерфейс стремится поддерживать распространенные языки программирования и фреймворки, включая возможность загрузки и установки новых языков и фреймворков. Цель интерфейса — собирать и извлекать данные с помощью кода. Такие системы, как Apache Spark, иллюстрируют потенциал использования различных языков, включая SQL, Java, Scala, Python и R, для обработки данных. Они напрямую вдохновляют AIDE.

Code Interface позволяет «опытным пользователям», знакомым с программированием, писать код на предпочитаемом ими языке без вмешательства ИИ, который может попытаться «оптимизировать» или изменить путь выполнения.

Например, пользователь может написать задание по манипулированию данными на Python для обработки большого набора данных, а затем переключиться на SQL, чтобы выполнить некоторые запросы к обработанным данным. Прямое выполнение кода гарантирует, что механизм выполнения Code Interface обработает код именно так, как он написан. Code Interface всегда становится доступным для любого устройства, которое позволяет вводить текст (или голос) или поддерживает низкую или отсутствующую графику кода.

Цель Universal Data Exchange — быть отзывчивым и интуитивно понятным, независимо от сложности задачи или предпочитаемого пользователем способа взаимодействия. Гибкость в модальностях гарантирует, что система не только удобна для пользователя и инклюзивна, но и способна поддерживать сложные, контекстно-зависимые рабочие процессы.

3.2.3. Искусственный интеллект и поток данных

Слой ИИ в AIDE состоит из нескольких компонентов ИИ. ИИ функционирует как инструмент для дополнения управления данными человека, а не как его замена. Любая реализация системы остается в соответствии с человеческими ценностями, целями и режимами регулирования. Цель состоит в том, чтобы построить слой ИИ на основе надежных, современных инженерных принципов [29] .

Работа слоя искусственного интеллекта заключается в сборе, хранении, обработке, организации данных и извлечении информации из данных для поддержки принятия решений и действий владельца. По сути, он соединяет владельца системы через универсальный обмен данными с уровнем хранения. Уровень ИИ взаимодействует с модальностями универсального обмена данными. Эти модальности, в свою очередь, могут передавать ИИ любые внешние источники, предоставленные владельцем (например, видеофайл).

Среди основных компонентов ИИ **генеральный менеджер** выполняет функции центрального оркестратора, контролируя и координируя действия нескольких специализированных агентов ИИ для достижения сложных, высокоуровневых целей. Действуя как мета-контролер, менеджер назначает задачи, отслеживает прогресс и динамически корректирует роли и приоритеты специализированных агентов на основе обратной связи в реальном времени и изменяющихся условий.

Функции генерального директора используют такие технологии искусственного интеллекта, как иерархическое обучение с подкреплением (позволяет автономно разбивать долгосрочные задачи на более простые подзадачи) [134] и метаобучение (облегчает сам процесс обучения) [181] .

Специализированные агенты ИИ предназначены для обработки определенных функций управления данными, таких как анализ данных, языковая обработка, поддержка принятия решений или автоматизация, в то время как агент генерального менеджера интегрирует их результаты, разрешает конфликты и обеспечивает соответствие общим целям и ценностям AIDE. Эта архитектура обеспечивает адаптивное, контекстно-зависимое принятие решений, где общий агент ИИ синтезирует идеи от различных агентов, выявляет пробелы или избыточность и организует скоординированные ответы. Такая система основана на исследованиях в области многоагентных систем и автономной координации [40, 76, 183] , где центральный интеллект управляет распределенной, специализированной на задачах и предметной области экспертизой.

Рассмотрим задачи для одного специализированного агента. Специализированный *агент хранения* призван постоянно обновлять оптимизированное (преобразованное) хранилище. Этот агент функционирует как посредник между необработанными

данными в исходном хранилище и слоем обработки ИИ, гарантируя, что данные остаются чистыми, структурированными и актуальными. Агент может использовать методы очистки данных, такие как обнаружение выбросов, импутация пропущенных значений и нормализация форматов данных для поддержания согласованности. Более того, он может использовать потоковые данные в реальном времени и стратегии пакетной обработки для обработки крупномасштабных обновлений данных.

Специализированному агенту ИИ также необходимо будет обрабатывать дополнение и слияние данных из нескольких источников. Связанные данные часто поступают из разнородных источников: сенсорных сетей, взаимодействий с пользователями, внешних баз данных и т. д. Агент будет отвечать за выравнивание и слияние этих разрозненных потоков данных, разрешение конфликтов и заполнение пробелов для создания единого, обогащенного набора данных. Исследования методов слияния данных (Dong & Srivastava, 2015) предлагают стратегии для устранения несоответствий и избыточности для повышения надежности данных и производительности ИИ в нисходящем направлении.

Агент хранения должен использовать передовые методы машинного обучения для динамической классификации и организации файлов, выходя за рамки традиционных иерархических структур папок. Используя контекстное понимание, ИИ может анализировать не только содержимое файлов, но и метаданные, контекст создания и шаблоны использования для создания значимых организационных структур. Например, система ИИ может распознавать, что финансовый отчет и прогноз продаж связаны через общие точки данных, и автоматически группировать их вместе в более широкой категории «Бизнес-стратегия». Это позволяет системе поддерживать более гибкую и адаптивную систему организации файлов, гарантируя, что файлы будет легче находить и логически связывать на основе реальных отношений, а не жестких путей к папкам.

Предиктивная категоризация позволяет системам хранения данных на основе ИИ предугадывать потребности пользователей, изучая индивидуальное и организационное поведение при управлении файлами. Например, если пользователь часто обращается к отчетам по анализу рынка в начале каждого квартала, система ИИ может заранее предлагать или организовывать эти файлы для легкого доступа. Семантическое индексирование еще больше расширяет эту возможность, понимая значение и связи между файлами, а не полагаясь исключительно на ключевые слова. Если пользователь ищет «взаимодействие с клиентами», ИИ может извлекать не только файлы, явно озаглавленные этими терминами, но и связанные документы, такие как отчеты об отзывах пользователей, журналы обслуживания клиентов и маркетинговые планы, на основе семантического сходства. Это более глубокое понимание обеспечивает более интуитивный и эффективный поиск и извлечение.

Еще один специализированный агент — Security. Агенту ИИ потребуется надежное управление и мониторинг для обеспечения целостности и безопасности оптимизированного хранилища. Агент будет реализовывать модели обнаружения аномалий для выявления и изоляции поврежденных данных, попыток несанкционированного доступа или сбоев системы. Например, система кибербезопасности ИИ может использовать агента для обнаружения необычных шаблонов сетевого трафика и соответствующего обновления правил брандмауэра. Способность агента поддерживать целостность данных будет определяться исследованиями в области безопасного машинного обучения (Papernot et al., 2018).

Система хранения данных на основе агентского ИИ также будет включать функции автономного управления для оптимизации операций и повышения надежности системы. Интеллектуальные механизмы резервного копирования и избыточности могут автоматически создавать и управлять резервными копиями на основе важности и

частоты использования файла. Например, файлы критически важных проектов можно было бы резервировать чаще и хранить в географически распределенных местах, чтобы свести к минимуму риск потери данных. Динамическая безопасность и управление доступом позволят ИИ корректировать разрешения файлов на основе контекстного понимания ролей и поведения пользователей. Если сотрудник финансового отдела попытается получить доступ к конфиденциальным файлам HR, ИИ может пометить попытку или автоматически ограничить доступ. Кроме того, функции оптимизации ресурсов позволят ИИ прогнозируемо управлять распределением хранилища, перенося редко используемые файлы в менее затратное хранилище, при этом сохраняя часто используемые файлы в высокопроизводительном хранилище.

3.2.4. Два ограждения – System Guardian и Direct Access

Для обеспечения автономности и контроля со стороны владельца системы AIDE реализует полунезависимого системного защитника. Он также позволяет полностью обойти ИИ через прямой доступ.

Системный **страж** — это специализированная организация, специально разработанная для защиты Владельца системы путем мониторинга и оценки поведения AIDE. Это независимый системный компонент, целью которого является не управление данными, а обеспечение работы системы в приемлемых моральных, правовых и этических рамках, которые ставят благополучие Владельца системы на первое место.

Этот Guardian — нейросимволическая система ИИ [22] . Она опирается на детерминированные, жестко запрограммированные правила с некоторым обучением и адаптивностью, управляемыми ИИ. Детерминированные принципы обеспечивают стабильную основу для соблюдения ключевых этических границ, ограничений безопасности и правил эксплуатации. Это обеспечивает прозрачное и предсказуемое поведение, направленное на продвижение интересов Владельца системы.

Между тем, компонент ИИ вводит адаптивный интеллект, позволяя Guardian выявлять возникающие риски, интерпретировать сложные сценарии и рекомендовать корректировки. Компонент ИИ в Guardian должен быть сведен к минимуму — достаточному для обеспечения гибкого и адаптивного поведения, но недостаточному для внесения нежелательной неопределенности или непрозрачности логики принятия решений, которые возникают при сильной зависимости от ИИ.

System Guardian заключается в посредничестве между внутренними процессами системы и ценностями, приоритетами и целями System Owner. Опираясь на исследования в области согласования ценностей и этики ИИ [31, 147] , Guardian отслеживает потенциальные конфликты или риски и накладывает ограничения на потенциально небезопасные операции уровня ИИ AIDE. Например, если действие ИИ может непреднамеренно поставить под угрозу конфиденциальность пользователя (например, путем обмена личной информацией с внешней системой без явного согласия владельца), Guardian вмешивается и останавливает действие, одновременно изменяя System Owner. Его детерминированное ядро определяет фундаментальные ограничения, такие как обеспечение целостности данных или предотвращение вредных результатов, в то время как компонент, управляемый ИИ, допускает гибкие ответы в тонких контекстах. Эта гибридная конструкция гарантирует, что Guardian остается как заслуживающим доверия, так и отзывчивым, в конечном итоге защищая автономию и благополучие System Owner.

AIDE реализует всегда доступный **прямой доступ** к данным, возможность обойти ИИ, предоставляя канал от универсального интерфейса данных к базовому хранилищу.

Это включает доступ к данным в их исходном необработанном формате. Это обеспечивает прозрачность, контроль и надежность.

Как и в случае с пользовательскими и кодовыми интерфейсами, прямой доступ гарантирует, что пользователи не зависят исключительно от ИИ для управления данными, что может вносить предвзятость, искажения или фильтрацию на основе данных обучения модели или внутренней эвристики. Получая доступ к необработанным данным напрямую, пользователи могут проверять сгенерированные ИИ идеи, обнаруживать несоответствия и применять независимые аналитические методы, когда результаты ИИ кажутся ненадежными или неясными. Такой подход повышает автономность пользователя и способность принимать решения, включая резервный механизм, который не опосредован ИИ.

System Guardian вместе с ядром AI имеют видимость данных, перемещаемых через Direct Access, но не могут вмешиваться в них (невмешательство обеспечивается неизменными правилами). Доступ Guardian к обходным действиям System Owner необходим Guardian для изучения шаблонов прямых действий пользователя. Эти изученные шаблоны могут быть впоследствии реализованы Guardian полуавтоматически, что экономит будущие усилия владельца на ручное вмешательство. Агенты AI также имеют видимость Direct Access, но это делается в целях обучения для улучшения управления данными.

Прямой доступ автоматически реализуется при использовании кода и пользовательского интерфейса и является опцией в основном мультимодальном интерфейсе.

3.2.5. Уровень хранения данных

Storage Layer — это место, где хранятся данные, которые должны сохраняться после сеанса в реальном времени. Он состоит из трех основных компонентов: исходные данные, оптимизированные (преобразованные) данные и контрольные данные.

Исходные, необработанные данные всегда хранятся в **Исходном контейнере** . Хранение данных в исходном, необработанном формате имеет важное значение для обеспечения целостности данных, воспроизводимости и гибкости в будущем анализе. Когда данные сохраняются в необработанном состоянии, они остаются свободными от потенциальных искажений, вносимых предварительной обработкой, преобразованием или интерпретацией с помощью ИИ, что позволяет проводить независимую проверку и повторную обработку по мере необходимости. Сохранение необработанных данных обеспечивает прослеживаемость, позволяя Владельцу системы или доверенной стороне проверять этапы обработки и сверять результаты с исходными данными. Более того, в таких областях, как научные исследования и финансовое моделирование, сохранение исходных данных имеет решающее значение для воспроизводимости и обеспечения возможности независимой проверки результатов [68] .

Хотя исходные данные не могут быть изменены, их следует сжимать до тех пор, пока не будет потерян сигнал (так называемое сжатие без потерь). Сжатие не должно использовать никаких фирменных алгоритмов, чтобы гарантировать, что данные всегда могут быть распакованы с помощью программного обеспечения с открытым исходным кодом, свободно доступного. Также рекомендуется хранить алгоритм сжатия в Data Controller Container (описано ниже).

Оптимизированный **контейнер данных** сохраняет выходные данные процессов предварительной обработки, очистки и преобразования, предназначенных для структурирования и уточнения необработанных данных в формат, который является одновременно значимым и эффективным для обработки ИИ. Этот уровень гарантирует, что данные свободны от шума, несоответствий и пропущенных значений.

Например, в обработке естественного языка (NLP) токенизация, лемматизация и удаление стоп-слов являются важными этапами предварительной обработки, которые повышают способность модели ИИ понимать и генерировать текст, похожий на человеческий. Аналогично, в моделях машинного обучения нормализация данных и кодирование категориальных переменных гарантируют, что числовые входные данные масштабируются соответствующим образом, предотвращая непропорциональное влияние на модели больших или малых значений [55, 90, 148] . Поскольку очистка и подготовка данных могут занимать до 80% времени, важно оптимизировать данные для извлечения. Без этого уровня системам ИИ было бы трудно обрабатывать необработанные, неструктурированные данные, что приводит к увеличению задержки, плохой производительности модели и снижению точности прогнозирования.

Контейнер **контроллера данных** — это расширенный репозиторий метаданных. Он служит централизованным местом хранения, которое управляет и организует критически важные метаданные, версионирование и информацию о сохраненных необработанных и обработанных данных. Он функционирует как центр управления, поддерживая проверяемую запись преобразований данных, журналов доступа и алгоритмических процессов, применяемых к данным. Это включает в себя хранение шаблонов проектирования и основного кода, используемого при управлении данными, например алгоритмов сжатия, методов шифрования и стратегий индексации. Это гарантирует, что операции с данными не только отслеживаются, но и обратимы при необходимости. Контейнер также облегчает управление версиями, управляя историческими снимками данных и кода, используемого для их обработки, что позволяет выполнять откат в случае ошибок обработки. Объединяя эту информацию, контроллер данных улучшает управление данными и соответствие нормативным требованиям, таким как GDPR и HIPAA, гарантируя, что обработка данных соответствует установленным стандартам прозрачности и подотчетности.

4. Потенциальные проблемы и стратегии их смягчения

Хотя потенциал ИИ в управлении хранением данных значителен, необходимо решить несколько проблем, чтобы обеспечить его успешное использование. Одной из главных проблем является *конфиденциальность* . Системы хранения данных на основе ИИ предназначены для обработки огромных объемов конфиденциальных данных, что вызывает опасения по поводу несанкционированного доступа и утечек данных. Чтобы снизить эти риски, AIDE должна реализовать надежные методы шифрования и требовать явного согласия пользователя перед обработкой или хранением конфиденциальной информации. Кроме того, дифференциальные методы конфиденциальности и механизмы контроля доступа могут повысить безопасность данных.

Еще одной проблемой является *алгоритмическая предвзятость* , которая может привести к несправедливому или неоптимальному принятию решений при управлении хранилищем на основе ИИ. Предвзятость может возникать из-за данных обучения, которым не хватает репрезентативности [117, 127, 128] , или алгоритмов, которые усиливают существующие предрассудки [35, 59, 61, 117, 130] . Чтобы противостоять этому, системы ИИ должны быть тщательно предварительно обучены с использованием разнообразных наборов данных, которые отражают различные потребности пользователей. Необходимы улучшенные методы смягчения предвзятости с акцентом на данные [23, 27, 71, 97, 149, 182] и алгоритмы [5, 23, 27, 62, 117, 170] . Примечательно, что предвзятость ИИ иногда возникает в базовой реальности, которая является дискриминационной и предвзятой [161] . Следовательно, организационные и психологические исследования причин, выявления и смягчения организационных и

личных предубеждений [11, 32, 63, 79, 80, 172] также могут способствовать техническим разработкам AIDE.

Надежность системы является еще одним критическим фактором, поскольку решения для хранения на основе ИИ должны функционировать бесперебойно, чтобы предотвратить потерю данных или сбои в работе. Без адекватных мер безопасности неожиданные ошибки или сбои ИИ могут привести к значительному снижению производительности. Чтобы решить эту проблему, внедрение резервных механизмов, таких как резервное копирование избыточной системы и варианты человеческого контроля, может обеспечить непрерывность. Регулярный мониторинг системы, обнаружение ошибок и отказоустойчивые протоколы имеют важное значение для повышения устойчивости хранилища на основе ИИ. В целом, как говорит Бостром [30] и другие утверждают [26, 58, 66, 78, 143, 171] , что ИИ должен основываться на прочных инженерных и моральных принципах.

Важным аспектом надежности является прозрачность логики принятия решений, используемой ИИ. Мы уже отмечали проблемы прозрачности, которые резко обострились с введением нейронных сетей глубокого обучения и больших языковых моделей. Существует много открытых вопросов, связанных с прозрачностью, таких как обратная разработка нейронных сетей [4, 129] и объяснимость больших языковых моделей [39, 114, 121, 186] . Многообещающей возможностью являются подходы, инвариантные к моделям; они стали особенно важными по мере роста сложности современных моделей ИИ [4, 13, 108, 139] .

Семантическая прозрачность , гарантирующая, что агенты ИИ полностью понимают значение данных, которыми они управляют, остается проблемой, несмотря на впечатляющий прогресс в области ИИ. Современные инструменты поиска и организации на основе ИИ ограничены в своей способности понимать контекст и нюансы различных типов данных, что приводит к неполным или нерелевантным результатам поиска [88, 95, 175] . *Галлюцинация* , склонность ИИ создавать бессмысленные данные, остается проблемой [34, 77, 185] . Некоторые утверждают, что проблемы семантики никогда не будут полностью решены, поскольку человеческий смысл не является сетью статистических вероятностей [19, 48, 177] . Как утверждают Хомский и коллеги, «системы машинного обучения могут узнать как то, что Земля плоская, так и то, что Земля круглая. Они просто торгуют вероятностями, которые со временем меняются. По этой причине прогнозы систем машинного обучения всегда будут поверхностными и сомнительными» [48] . Не все разделяют эту пессимистическую точку зрения. Некоторые утверждают, что ИИ не обязательно должен думать как человек, чтобы быть полезным [24] . В более общем плане, исследования семантики ИИ продолжают продвигать прогресс [34, 77, 173, 185] . Такие области, как *большие модели рассуждений,* имеют особые перспективы, поскольку они сочетают статистические вероятности с логическими рассуждениями [136] . Большие модели действий, ключевая технология агентного ИИ [28, 45] ,

Еще одной ключевой проблемой является *совместимость данных и систем* . Хранилище на основе ИИ должно легко интегрироваться с различными существующими платформами, форматами файлов и корпоративным программным обеспечением. Проблемы несовместимости могут привести к неэффективности, разрозненности данных и сбоям в работе. Чтобы смягчить это, AIDE следует принять стандартизированные протоколы обмена данными и решения на основе API для упрощения интеграции в различных цифровых экосистемах. Обеспечение кроссплатформенной совместимости и масштабируемости имеет решающее значение для максимизации преимуществ управления хранилищем на основе ИИ.

Другой важной проблемой является координация нескольких систем AIDE, каждая из которых принадлежит своим владельцам. Такая потребность будет очевидна в

организационных условиях, где сотрудники и отделы могут иметь свои собственные AIDE. Эту задачу можно решить с помощью текущих исследований по координации многоагентных систем [40, 76, 183] . Разработка высокоуровневых координирующих систем, специально предназначенных для координации нескольких AIDE, является еще одной возможностью.

При условии, что AIDE будет широко использоваться, *экологическая устойчивость* остается важной проблемой. Хотя целью AIDE является снижение воздействия на окружающую среду, что, как мы ожидаем, произойдет из-за оптимизации деятельности по управлению данными, можно ожидать эффекта отскока. Эффект отскока или возврата возникает, когда рост эффективности приводит к увеличению потребления, потенциально компенсируя первоначальную экономию [169] . Следовательно, экономия энергии и снижение воздействия на окружающую среду всегда должны быть движущейся целью, в идеале предотвращая любые эффекты отскока. Одним из долгосрочных решений, направленных на управление зелеными данными, является *принципиально зеленая технология* . Это технология, которая даже в больших масштабах не загрязняет окружающую среду. Например, люди используют воздух для вербального общения. Похоже, что это не оказывает заметного воздействия на окружающую среду.

Квантовые вычисления, основанные на самой распространенной субстанции во Вселенной, элементарных частицах, обещают в будущем обеспечить принципиально зеленую энергию [106] . Однако для этого необходимо устранить существенную неэффективность текущих квантовых решений [15, 96, 137, 146] .

Ограничения возможностей устройств представляют собой проблему для AIDE, поскольку точность, разрешение и объем сбора и представления данных по своей сути связаны с физическими и техническими ограничениями устройств. Например, дроны с ограниченным временем полета из-за емкости аккумулятора могут не собирать исчерпывающие данные на больших площадях, а тактильные интерфейсы могут не иметь точности, необходимой для подробной обратной связи. Более того, вычислительная мощность и диапазон датчиков смартфонов и ноутбуков могут ограничивать сложность анализа в реальном времени. Основная проблема и возможность — предоставление интерфейсов для в настоящее время недостаточно используемых человеческих чувств, таких как обоняние, вкус и тактильные ощущения [18, 126, 174] . Большая часть реальности недоступна для людей из-за их сенсорных ограничений (способных общаться в узкой полосе аудиовизуальных спектров). Знаменитый биолог EO Wilson [179] называет людей «сенсорными идиотами», поскольку 99% видов имеют значительно превосходящие сенсорные системы. Обонятельные, вкусовые и тактильные интерфейсы могут позволить людям заново открыть для себя мир, в котором они живут. Еще одной проблемой является интеграция сенсорных модальностей в единый, универсальный интерфейс.

Этические соображения также играют решающую роль в принятии ИИ для управления хранением данных. Такие вопросы, как владение данными, подотчетность и прозрачность, должны быть хорошо реализованы, чтобы способствовать доверию между пользователями. Установление четких руководящих принципов относительно обязанностей ИИ, обеспечение контроля пользователей над своими данными и поддержание открытого общения о том, как ИИ принимает решения, может помочь смягчить этические проблемы. Соблюдение нормативных требований законов о защите данных также должно быть приоритетом. Продвижение исследований в области общей этики (которая является ювенальной областью [138]) и этики в ИИ [26, 58, 66, 78, 143, 171] имеет первостепенное значение для того, чтобы AIDE действительно ставила людей на первое место во всех своих решениях.

Еще одной проблемой являются *угрозы кибербезопасности* . Системы ИИ, обрабатывающие большие объемы конфиденциальных данных, могут стать основными целями для кибератак, включая программы-вымогатели, отравление данных и угрозы враждебного ИИ. Чтобы противостоять этим рискам, организации должны инвестировать в передовые механизмы обнаружения угроз, внедрять многоуровневые архитектуры безопасности и проводить регулярные аудиты безопасности. Крайне важно оставаться далеко впереди угроз кибербезопасности и предвосхищать развитие событий, насколько это возможно. Преступники уже собирают персональные данные в надежде, что более умный ИИ и квантовые компьютеры позволят расшифровывать данные через десятилетия [157] . Обеспечение будущего и опережение развития кибербезопасности являются первостепенной задачей.

Масштабируемость представляет собой еще одно существенное препятствие. По мере роста потребностей в хранении данных решения ИИ должны иметь возможность масштабироваться без снижения производительности. Обеспечение того, чтобы алгоритмы ИИ могли эффективно справляться с возросшими рабочими нагрузками и разнообразными структурами данных, имеет решающее значение. Реализация модульной и гибкой архитектуры ИИ, которая допускает инкрементные обновления и оптимизацию производительности, поможет поддерживать масштабируемость, не отставая от меняющихся требований к хранению. Здесь снова критически важна готовность к будущему любых решений для хранения, поскольку объемы данных могут внезапно резко возрасти, требуя немедленных решений. Особые перспективы открывает работа над масштабируемым и распределенным хранилищем [57, 163] .

Организации и регулирующие органы также должны ориентироваться в *экономических и организационных последствиях* внедрения AIDE. Системы хранения на основе ИИ обладают потенциалом значительного сокращения времени, затрачиваемого на управление данными. Такое повышение эффективности позволяет сотрудникам сосредоточиться на более ценной работе, повышая общую производительность. Тем не менее, вопросы смещения рабочих мест, доверия, интеграции в существующие организационные процедуры, последствия для организационной стратегии и общей динамики рынка должны быть тщательно изучены для устранения любых негативных последствий.

Проактивно решая эти проблемы и стратегически используя возможности ИИ, организации и отдельные лица могут раскрыть весь потенциал агентного ИИ в управлении хранилищами.

5. Заключение

Генерация цифровых данных в последние годы выросла в геометрической прогрессии, что создало беспрецедентные проблемы в хранении, организации и поиске. Традиционные системы управления данными в значительной степени зависят от ручного вмешательства человека, что приводит к неэффективности, проблемам доступности и значительным рискам. Искусственный интеллект, основанный на таких прорывах, как глубокое обучение, большие языковые модели, агентный ИИ, представляет собой решение, которое может устранить фундаментальные ограничения управления данными, внедряя интеллектуальные, автономные возможности управления данными в масштабе и для всех.

Агентный ИИ представляет собой смену парадигмы в цифровом хранении, выходящую за рамки традиционных парадигм управления файлами. Внедряя интеллектуальные автономные системы, способные понимать контекст, предсказывать потребности пользователей и динамически управлять цифровыми активами, мы можем

создать более эффективное, безопасное и удобное для пользователя управление данными для всех.

6. Ссылки

1. Aaltonen, A. et al.: What is Missing from Research on Data in Information Systems? Insights from the Inaugural Workshop on Data Research. Communications of the Association for Information Systems. 53, 1, 17 (2023).
2. Aaltonen, A., Stelmaszak, M.: Data Innovation Lens: A New Way to Approach Data Design as Value Creation. Available at SSRN 4574855. (2024).
3. Abbasi, A. et al.: Text Analytics to Support Sense-Making in Social Media: A Language-Action Perspective. MIS Quarterly. 42, 2, 1–38 (2018).
4. Adadi, A., Berrada, M.: Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE access. 6, 52138–52160 (2018).
5. Adebayo, J.A.: FairML: ToolBox for diagnosing bias in predictive modeling. (2016).
6. Agichtein, E. et al.: Finding high-quality content in social media. In: Proceedings of the international conference on Web search and web data mining. pp. 183–194 ACM, Palo Alto, California, USA (2008).
7. Alaimo, C., Kallinikos, J.: Data rules: Reinventing the market economy. MIT Press, Cambridge, MA (2024).
8. Alavi, M. et al.: A Knowledge Management Perspective of Generative Artificial Intelligence. Journal of the Association for Information Systems. 25, 1, 1–12 (2024).
9. Anderson, J.R.: Cognitive psychology and its implications. WH Freeman/Times Books/Henry Holt & Co (1985).
10. Arazy, O. et al.: A theory-driven design framework for social recommender systems. Journal of the Association for Information Systems. 11, 9, 455–490 (2010).
11. Archambault, A. et al.: A natural bias for the basic level? Twenty-Second Annual Conference of the Cognitive Science Society. 585-590 1075 (2000).
12. Arora, L. et al.: A comprehensive review on NUI, multi-sensory interfaces and UX design for applications and devices for visually impaired users. Frontiers in Public Health. 12, 1357160 (2024).
13. Arrieta, A.B. et al.: Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion. 58, 82–115 (2020).
14. Baird, A., Maruping, L.M.: The Next Generation of Research on IS Use: A Theoretical Framework of Delegation to and from Agentic IS Artifacts. MIS quarterly. 45, 1, (2021).
15. Barad, K.: Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning. Duke University Press, Durham, North Carolina (2007).
16. Batini, C., Scannapieca, M.: Data quality: concepts, methodologies and techniques. Springer (2006).
17. Beck, K. et al.: Manifesto for agile software development. Agile Alliance. (2001).
18. Benali-Khoudja, M. et al.: Tactile interfaces: a state-of-the-art survey. Presented at the Int. symposium on robotics (2004).
19. Bender, E.M. et al.: On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? Presented at the Proceedings of the 2021 ACM conference on fairness, accountability, and transparency (2021).
20. Bernstein, D.J., Lange, T.: Post-quantum cryptography. Nature. 549, 7671, 188–194 (2017).

21. von Bertalanffy, L.: General system theory: Foundations, development, applications. Braziller, New York. (1968).
22. Bhuyan, B.P. et al.: Neuro-symbolic artificial intelligence: a survey. Neural Computing and Applications. 36, 21, 12809–12844 (2024).
23. Bird, T.J. et al.: Statistical solutions for error and bias in global citizen science datasets. Biological Conservation. 173, 144–154 (2014).
24. Bolhuis, J.J. et al.: Three reasons why AI doesn't model human language. Nature. 627, 8004, 489–489 (2024).
25. Bonnie, E.: 110+ of the Latest Data Breach Statistics [Updated 2025], https://secureframe.com/blog/data-breach-statistics, last accessed 2025/04/01.
26. Borenstein, J., Howard, A.: Emerging challenges in AI and the need for AI ethics education. AI and Ethics. 1, 1, 61–65 (2021).
27. Borisov, V. et al.: Deep neural networks and tabular data: A survey. IEEE Transactions on Neural Networks and Learning Systems. (2022).
28. Bornet, P. et al.: Agentic Artificial Intelligence: Harnessing AI Agents to Reinvent Business, Work and Life. Irreplaceable Publishing, NA (2025).
29. Bostrom, N.: How long before superintelligence? International Journal of Futures Studies. 2, 1, 1–9 (1998).
30. Bostrom, N.: Superintelligence: Paths, Dangers, Strategies. Oxford University Press, Oxford, England (2014).
31. Bostrom, N., Yudkowsky, E.: The ethics of artificial intelligence. In: Artificial intelligence safety and security. pp. 57–69 Chapman and Hall/CRC (2018).
32. Brewer, M.B.: In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. Psychological Bulletin. 86, 2, 307–324 (1979).
33. Briney, K.: Data Management for Researchers: Organize, maintain and share your data for research success. Pelagic Publishing Ltd, Exeter, UK (2015).
34. Buchanan, J. et al.: ChatGPT Hallucinates Non-existent Citations: Evidence from Economics. The American Economist. 69, 1, 80–87 (2024).
35. Buranyi, S.: Rise of the racist robots–how AI is learning all our worst impulses. The Guardian. August. 8, (2017).
36. Burton-Jones, A., Volkoff, O.: How can we develop contextualized theories of effective use? A demonstration in the context of community-care electronic health records. Information Systems Research. 28, 3, 468–489 (2017).
37. Byrne, E. et al.: Cyberbullying and social media: Information and interventions for school nurses working with victims, students, and families. The Journal of School Nursing. 34, 1, 38–50 (2018).
38. Cabral, J.: Is generation Y addicted to social media? The Elon Journal of Undergraduate Research in Communications. 2, 1, 5–14 (2011).
39. Cambria, E. et al.: Xai meets llms: A survey of the relation between explainable ai and large language models. arXiv preprint arXiv:2407.15248. (2024).
40. Cao, Y. et al.: An overview of recent progress in the study of distributed multi-agent coordination. IEEE Transactions on Industrial informatics. 9, 1, 427–438 (2012).
41. Castellanos, A. et al.: Basic Classes in Conceptual Modeling: Theory and Practical Guidelines. Journal of the Association for Information Systems. 21, 4, 1001–1044 (2020).
42. Castellanos, A. et al.: Understanding Benefits and Limitations of Unstructured Data Collection for Repurposing Organizational Data. In: Wrycza, S. and Maślankowski, J. (eds.) Information Systems: Research, Development, Applications, Education: 10th SIGSAND/PLAIS EuroSymposium 2017, Gdansk, Poland, September 22, 2017, Proceedings. pp. 13–24 Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-66996-0_2.

43. Chandra Kruse, L. et al.: Making Use of Design Principles. In: LNCS 9661. pp. 37–51 Springer, Berlin / Heidelberg (2016).
44. Chandra Kruse, L., Seidel, S.: Tensions in Design Principle Formulation and Reuse. In: DESRIST 2017. (2017).
45. Chang, Y. et al.: A survey on evaluation of large language models. ACM Transactions on Intelligent Systems and Technology. (2023).
46. Chen, H. et al.: Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly. 36, 4, 1165–1188 (2012).
47. Chipidza, W. et al.: Quantum Computing and IS-Harnessing the Opportunities of Emerging Technologies. Communications of the Association for Information Systems. 52, 1, 7 (2023).
48. Chomsky, N. et al.: Opinion | Noam Chomsky: The False Promise of ChatGPT, https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html, (2023).
49. Chua, C.E.H. et al.: Data Management. MIS Quarterly Online. 1–10 (2022).
50. Church, R.M.: The effective use of secondary data. Learning and motivation. 33, 1, 32–45 (2002).
51. Combi, C. et al.: A modular approach to the specification and management of time duration constraints in BPMN. Information Systems. 84, 111–144 (2019).
52. Cooper, C.: Citizen science: How ordinary people are changing the face of discovery. Abrams, New York NY (2016).
53. DAMA et al.: DAMA-DMBOK: Data Management Body of Knowledge. Technics Publications, Sedona, AZ (2024).
54. Davenport, T. et al.: How artificial intelligence will change the future of marketing. Journal of the Academy of Marketing Science. 48, 24–42 (2020).
55. Duboue, P.: The Art of Feature Engineering: Essentials for Machine Learning. Cambridge University Press, Cambridge, UK (2020).
56. Eriksson, O. et al.: The case for classes and instances-a response to representing instances: the case for reengineering conceptual modelling grammars. European Journal of Information Systems. 28, 6, 681–693 (2019).
57. Faniel, I.M., Zimmerman, A.: Beyond the data deluge: A research agenda for large-scale data sharing and reuse. International Journal of Digital Curation. 6, 1, 58–69 (2011).
58. Floridi, L., Cowls, J.: A unified framework of five principles for AI in society. In: Ethics, Governance, and Policies in Artificial Intelligence. pp. 5–17 Springer (2021).
59. Friedman, B., Nissenbaum, H.: Bias in computer systems. ACM Transactions on information systems (TOIS). 14, 3, 330–347 (1996).
60. Gabernet, A.R., Limburn, J.: Breaking the 80/20 rule: How data catalogs transform data scientists' productivity, https://www.ibm.com/cloud/blog/ibm-data-catalog-data-scientists-productivity, last accessed 2023/01/01.
61. Garcia, M.: Racist in the machine: The disturbing implications of algorithmic bias. World Policy Journal. 33, 4, 111–117 (2016).
62. George, J.F. et al.: Countering the anchoring and adjustment bias with decision support systems. Decision Support Systems. 29, 2, 195–206 (2000).
63. Gilovich, T. et al.: Heuristics and biases: The psychology of intuitive judgment. Cambridge University Press, Cambridge, UK (2002).
64. Grande, D. et al.: Reducing data costs without sacrificing growth | McKinsey, (2020).
65. Gregor, S., Hovorka, D.: Causality: The elephant in the room in information systems epistemology. In: 19th European Conference on Information Systems. pp. 1–10 , Helsinki, Finland (2011).
66. Hagendorff, T.: The ethics of AI ethics: An evaluation of guidelines. Minds and Machines. 30, 1, 99–120 (2020).

67. Harmon, R.: Admissibility standards for scientific evidence. In: Microbial Forensics. pp. 381–392 Elsevier (2005).
68. Hevner, A. et al.: Transparency in Design Science Research. DSS. 182, 1, 1–19 (2024).
69. Horton, E.L. et al.: A review of principles in design and usability testing of tactile technology for individuals with visual impairments. Assistive technology. 29, 1, 28–36 (2017).
70. Hovorka, D.S., Germonprez, M.: Tinkering, tailoring and bricolage: Implications for theories of design. Proceedings of the 15th Americas Conference on Information Systems. (2009).
71. Hufnagel, E.M., Conca, C.: User Response Data: The Potential for Errors and Biases. Information Systems Research. 5, 1, 48–73 (1994).
72. Hvalshagen, M. et al.: Empowering Users with Narratives: Examining The Efficacy Of Narratives For Understanding Data-Oriented Conceptual Models. Information Systems Research. 34, 3, 890–909 (2023).
73. IEA: Data centres & networks, https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks, last accessed 2024/09/13.
74. Ipeirotis, P.G. et al.: Quality management on amazon mechanical turk. In: Proceedings of the ACM SIGKDD workshop on human computation. pp. 64–67 ACM (2010).
75. Ixmeier, A. et al.: Green Is Everywhere And Nowhere Should Information Systems Scholarship Engage in Sustainability? In: ECIS 2024. (2024).
76. Jennings, N.R.: Commitments and conventions: The foundation of coordination in multi-agent systems. The knowledge engineering review. 8, 3, 223–250 (1993).
77. Ji, Z. et al.: Survey of hallucination in natural language generation. ACM Computing Surveys. 55, 12, 1–38 (2023).
78. Jobin, A. et al.: The global landscape of AI ethics guidelines. Nature Machine Intelligence. 1, 9, 389–399 (2019).
79. Jussim, L. et al.: Prejudice, stereotypes, and labeling effects: Sources of bias in person perception. Journal of personality and social psychology. 68, 2, 228–246 (1995).
80. Kahneman, D.: Thinking, fast and slow. Macmillan Publishing Co., London, UK (2011).
81. Kar, A.K. et al.: How could quantum computing shape information systems research–An editorial perspective and future research directions. International Journal of Information Management. 102776 (2024).
82. Khan, A., Khusro, S.: An insight into smartphone-based assistive solutions for visually impaired and blind people: issues, challenges and opportunities. Universal Access in the Information Society. 20, 2, 265–298 (2021).
83. Kim, H.K. et al.: The interaction experiences of visually impaired people with assistive technology: A case study of smartphones. International Journal of Industrial Ergonomics. 55, 22–33 (2016).
84. Kim, H.N. et al.: Accessible haptic user interface design approach for users with visual impairments. Universal access in the information society. 13, 415–437 (2014).
85. Kitchens, B. et al.: Understanding Echo Chambers and Filter Bubbles: The Impact of Social Media on Diversification and Partisan Shifts in News Consumption. MIS Quarterly. 44, 4, (2020).
86. Kurtz, J.E., Pintarelli, E.M.: The Daubert standards for admissibility of evidence based on the Personality Assessment Inventory. Psychological Injury and Law. 17, 2, 105–116 (2024).
87. Kushner, A.M., Guan, Z.: Modular design in natural and biomimetic soft materials. Angewandte Chemie International Edition. 50, 39, 9026–9057 (2011).

88. Larsen, K.R. et al.: Understanding the Elephant: The Discourse Approach to Boundary Identification and Corpus Construction for Theory Review Articles. Journal of the Association for Information Systems. 20, 7, 15 (2019).
89. Larsen, K.R. et al.: Validity in Design Science. MIS Quarterly. 1–40 (2025).
90. Larsen, K.R., Becker, D.S.: Automated Machine Learning for Business:An Introduction to Accurate, Easy, and Fast Analytics. Oxford, UK, New York NY (2020).
91. Lavie, N. et al.: Load theory of selective attention and cognitive control. Journal of Experimental Psychology: General. 133, 3, 339 (2004).
92. Law, M.: Reduce, reuse, recycle: Issues in the secondary use of research data. IASSIST Quarterly. 29, 1, 5–5 (2006).
93. Lee, Y.W.: Crafting Rules: Context-Reflective Data Quality Problem Solving. Journal of Management Information Systems. 20, 3, 93–119 (2003).
94. Leroux, O.: Legal admissibility of electronic evidence. International Review of Law, Computers & Technology. 18, 2, 193–220 (2004).
95. Li, J. et al.: TheoryOn: Designing a Construct-Based Search Engine to Reduce Information Overload for Behavioral Science Research. MIS quarterly. 44, 4, 1733–1772 (2020).
96. Lidar, D.A., Brun, T.A.: Quantum error correction. Cambridge university press, Cambridge (2013).
97. Lohr, S.L.: Sampling: Design and Analysis. Cengage Learning (2009).
98. Loos, P. et al.: Green IT: a matter of business and information systems engineering? Business & Information Systems Engineering. 3, 4, 245–252 (2011).
99. Lukyanenko, R. et al.: Artifact Sampling: Using Multiple Information Technology Artifacts to Increase Research Rigor. In: Proceedings of the 51st Hawaii International Conference on System Sciences (HICSS 2018). pp. 1–12 , Big Island, Hawaii (2018).
100. Lukyanenko, R. et al.: Datish: A Universal Conceptual Modeling Language to Model Anything by Anyone. In: ER Forum 2023. pp. 1–14 Springer, Lisbon, Portugal (2023).
101. Lukyanenko, R. et al.: Expecting the Unexpected: Effects of Data Collection Design Choices on the Quality of Crowdsourced User-generated Content. MISQ. 43, 2, 634–647 (2019).
102. Lukyanenko, R. et al.: Guidelines for Establishing Instantiation Validity in IT Artifacts: A Survey of IS Research. In: DESRIST 2015, LNCS 9073. Springer, Berlin / Heidelberg (2015).
103. Lukyanenko, R. et al.: Inclusive Conceptual Modeling: Diversity, Equity, Involvement, and Belonging in Conceptual Modeling. In: ER Forum 2023. pp. 1–4 Springer, Lisbon, Portugal (2023).
104. Lukyanenko, R.: Information Quality Research Challenge: Information Quality in the Age of Ubiquitous Digital Intermediation. J. Data and Information Quality. 7, 1–2, 3:1-3:3 (2016). https://doi.org/10.1145/2856038.
105. Lukyanenko, R. et al.: Instantiation Validity in IS Design Research. In: DESRIST 2014, LNCS 8463. pp. 321–328 Springer (2014).
106. Lukyanenko, R.: Quantum Data Management to Reduce Environmental Impact of IT. In: SIG GREEN 2024. pp. 1–5 , Bangkok, Thailand (2025).
107. Lukyanenko, R. et al.: Representing Instances: The Case for Reengineering Conceptual Modeling Grammars. European Journal of Information Systems. 28, 1, 68–90 (2019).
108. Lukyanenko, R. et al.: Superimposition: Augmenting Machine Learning Outputs with Conceptual Models for Explainable AI. In: 1st International Workshop on Conceptual Modeling Meets Artificial Intelligence and Data-Driven Decision Making. pp. 1–12 Springer, Vienna, Austria (2020).

109. Lukyanenko, R. et al.: The Impact of Conceptual Modeling on Dataset Completeness: A Field Experiment. In: Proceedings of the International Conference on Information Systems (ICIS). pp. 1–18 (2014).
110. Lukyanenko, R.: The MAGIC of Data Management: Understanding the Value and Activities of Data Management. arXiv preprint arXiv:2408.07607. (2024).
111. Lukyanenko, R. et al.: Universal Conceptual Modeling: Principles, Benefits, and an Agenda for Conceptual Modeling Research. Software & Systems Modeling. 1–35 (2024).
112. Lukyanenko, R., Parsons, J.: Design Theory Indeterminacy: What is it, how can it be reduced, and why did the polar bear drown? Journal of the Association for Information Systems. 21, 5, 1–30 (2020).
113. Lukyanenko, R., Weber, R.: A Realist Ontology of Digital Objects and Digitalized Systems. In: "Digital First" Era — A Joint AIS SIGSAND/SIGPrag Workshop. pp. 1–5 , Virtual Workshop (2022).
114. Luo, H., Specia, L.: From understanding to utilization: A survey on explainability for large language models. arXiv preprint arXiv:2401.12874. (2024).
115. Mayer-Schönberger, V., Cukier, K.: Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt (2013).
116. McAfee, A., Brynjolfsson, E.: Machine, platform, crowd: Harnessing our digital future. WW Norton & Company, New York, NY (2017).
117. Mehrabi, N. et al.: A survey on bias and fairness in machine learning. ACM Computing Surveys (CSUR). 54, 6, 1–35 (2021).
118. Metatla, O. et al.: Voice user interfaces in schools: Co-designing for inclusion with visually-impaired and sighted pupils. Presented at the Proceedings of the 2019 CHI conference on human factors in computing systems (2019).
119. Miller, G.: The magical number seven, plus or minus two: Some limits on our capacity for processing information. The psychological review. 63, 81–97 (1956).
120. Morgan, S.: Cybercrime To Cost The World $10.5 Trillion Annually By 2025, https://cybersecurityventures.com/hackerpocalypse-cybercrime-report-2016/, (2018).
121. Mumuni, F., Mumuni, A.: Explainable artificial intelligence (XAI): from inherent explainability to large language models. arXiv preprint arXiv:2501.09967. (2025).
122. Murphy, E. et al.: An empirical investigation into the difficulties experienced by visually impaired Internet users. Universal Access in the Information Society. 7, 79–91 (2008).
123. Murphy, H.: She Helped Crack the Golden State Killer Case. Here's What She's Going to Do Next., https://www.nytimes.com/2018/08/29/science/barbara-rae-venter-gsk.html, (2018).
124. Nelson, R.R. et al.: Antecedents of information and system quality: an empirical examination within the context of data warehousing. Journal of Management Information Systems. 21, 4, 199–235 (2005).
125. Norman, D.A.: The design of everyday things. Bsic Books, New York, NY (2002).
126. Obrist, M. et al.: Touch, taste, & smell user interfaces: The future of multisensory HCI. Presented at the Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems (2016).
127. Ogunseye, S. et al.: Do Crowds Go Stale? Exploring the Effects of Crowd Reuse on Data Diversity. In: WITS 2017. , Seoul, South Korea (2017).
128. Ogunseye, S., Parsons, J.: Can Expertise Impair the Quality of Crowdsourced Data? In: SIGOPEN Developmental Workshop at ICIS 2016. (2016).

129. Oh, S.J. et al.: Towards reverse-engineering black-box neural networks. Explainable AI: interpreting, explaining and visualizing deep learning. 121–144 (2019).
130. O'Neil, C.: Weapons of math destruction: How big data increases inequality and threatens democracy. Broadway Books, New York NY (2016).
131. Paas, F. et al.: Cognitive load theory and instructional design: Recent developments. Educational psychologist. 38, 1, 1–4 (2003).
132. Pang, M.-S. et al.: DIGITAL TECHNOLOGIES AND THE ADVANCEMENT OF SOCIAL JUSTICE: A FRAMEWORK AND AGENDA. MIS Quarterly. 48, 4, (2024).
133. Park, J. et al.: Green cloud? An empirical analysis of cloud computing and energy efficiency. Management Science. 69, 3, 1639–1664 (2023).
134. Pateria, S. et al.: Hierarchical reinforcement learning: A comprehensive survey. ACM Computing Surveys (CSUR). 54, 5, 1–35 (2021).
135. Pitron, G.: The Dark Cloud: how the digital world is costing the earth. Scribe Publications Pty Limited, London, UK (2023).
136. Plaat, A. et al.: Reasoning with large language models, a survey. arXiv preprint arXiv:2407.11511. (2024).
137. Preskill, J.: Quantum computing in the NISQ era and beyond. Quantum. 2, 79 (2018).
138. Rachels, J., Rachels, S.: The elements of moral philosophy 7e. McGraw Hill, New York NY (2012).
139. Rai, A.: Explainable AI: From black box to glass box. Journal of the Academy of Marketing Science. 48, 1, 137–141 (2020).
140. Redman, T.C.: Bad data costs the US $3 trillion per year. Harvard Business Review. 22, 11–18 (2016).
141. Redman, T.C.: Data quality for the information age. Artech House, Norwood, MA (1996).
142. Rietsche, R. et al.: Quantum computing. Electronic Markets. 32, 4, 2525–2536 (2022).
143. Robert Jr, L.P. et al.: Introduction to the special issue on AI fairness, trust, and ethics. AIS Transactions on Human-Computer Interaction. 12, 4, 172–178 (2020).
144. Rodriguez, J.: Some AI Lessons from Watson's Failure at MD Anderson, https://jrodthoughts.medium.com/some-ai-lessons-from-watsons-failure-at-md-anderson-9b895cf70840, last accessed 2024/07/28.
145. Roelens, B., Bork, D.: An evaluation of the intuitiveness of the PGA modeling language notation. In: Enterprise, Business-Process and Information Systems Modeling. pp. 395–410 Springer (2020).
146. Rosenblum, B., Kuttner, F.: Quantum enigma: Physics encounters consciousness. Oxford University Press (2011).
147. Russell, S.: Human compatible: AI and the problem of control. Penguin UK, London England (2019).
148. Russell, S.J., Norvig, P.: Artificial intelligence: a modern approach. Pearson Education Limited, Malaysia (2016).
149. Salk, C.F. et al.: Assessing quality of volunteer crowdsourcing contributions: lessons from the Cropland Capture game. International Journal of Digital Earth. 1, 1–17 (2015).
150. Samuel, B.M. et al.: Exploring the Effects of Extensional Versus Intentional Representations on Domain Understanding. MIS Quarterly. 42, 4, 1187–1209 (2018).
151. Sanderson, K.: AI science search engines explode in number. Nature. 616, 639–640 (2023).

152. Sarioglu, A. et al.: How Inclusive is Conceptual Modeling? A Systematic Review of Literature and Tools for Disability-aware Conceptual Modeling. In: 42nd International Conference on Conceptual Modeling (ER 2023). pp. 1–15 (2023).
153. Sernadas, C. et al.: Modular construction of logic knowledge bases: An algebraic approach. Information Systems. 15, 1, 37–59 (1990).
154. Sheng, V.S. et al.: Get another label? improving data quality and data mining using multiple, noisy labelers. In: ACM SIG KDD. pp. 614–622 (2008).
155. Shneiderman, B.: Designing the user interface. Pearson Education, Boston, MA (2003).
156. Shneiderman, B.: Human-centered AI. Oxford University Press, Oxford, UK (2022).
157. Singh, H.: Managing the Quantum Cybersecurity Threat: Harvest Now, Decrypt Later. In: Quantum Computing. pp. 142–158 CRC Press, New York NY (2024).
158. Skyttner, L.: General systems theory: origin and hallmarks. Kybernetes. 26, 6, 16–22 (1996).
159. Smith, C.U.M.: Biology of sensory systems. John Wiley & Sons, Hoboken, NJ (2008).
160. Someh, I. et al.: Ethical issues in big data analytics: A stakeholder perspective. Communications of the Association for Information Systems. 44, 1, 34 (2019).
161. Storey, V.C. et al.: Explainable AI: Opening the Black Box or Pandora's Box? Communications of the ACM. 66, 4, 1–6 (2022).
162. Storey, V.C. et al.: Generative Artificial Intelligence: Evolving Technology, Growing Societal Impact, and Opportunities for Research. Information Systems Frontiers. (2025).
163. Strengholt, P.: Data Management at scale. O'Reilly Media, Inc, New York (2020).
164. Strong, D.M. et al.: Data quality in context. Communications of the ACM. 40, 5, 103–110 (1997).
165. Sullivan, T., Adler, S.: Work, stress, and disability. International journal of law and psychiatry. 22, 5–6, 417–424 (1999).
166. Tams, S.: Helping older workers realize their full organizational potential: a moderated mediation model of age and IT-enabled task performance. MIS Q. 46, 1, 1–30 (2022).
167. Tams, S. et al.: Worker Stress in the Age of Mobile Technology: The Combined Effects of Perceived Interruption Overload and Worker Control. Journal of Strategic Information Systems. 1–40 (2020).
168. Tegarden, D. et al.: Systems Analysis and Design: An Object-Oriented Approach with UML. Wiley, Hoboken, NJ (2025).
169. Thiesen, J. et al.: Rebound effects of price differences. The International Journal of Life Cycle Assessment. 13, 104–114 (2008).
170. Tremblay, M.C. et al.: Using data mining techniques to discover bias patterns in missing data. Journal of Data and Information Quality (JDIQ). 2, 1, 2 (2010).
171. Tsamados, A. et al.: The ethics of algorithms: key problems and solutions. AI & SOCIETY. (2021). https://doi.org/10.1007/s00146-021-01154-8.
172. Tversky, A. et al.: Judgment under uncertainty: Heuristics and biases. Rationality in action: Contemporary approaches. 171–188 (1990).
173. Vaswani, A. et al.: Attention is all you need. Advances in neural information processing systems. 30, (2017).
174. Vi, C.T. et al.: Gustatory interface: The challenges of 'how'to stimulate the sense of taste. Presented at the Proceedings of the 2nd acm sigchi international workshop on multisensory approaches to human-food interaction (2017).

175. Wagner, G. et al.: Artificial intelligence and the conduct of literature reviews. Journal of Information Technology. 37, 2, 209–226 (2022).
176. Wand, Y., Wang, R.Y.: Anchoring data quality dimensions in ontological foundations. Communications of the ACM. 39, 11, 86–95 (1996).
177. Wang, A. et al.: Against predictive optimization: On the legitimacy of decision-making algorithms that optimize predictive accuracy. ACM Journal on Responsible Computing. (2023).
178. Wang, R.Y., Strong, D.M.: Beyond accuracy: what data quality means to data consumers. Journal of Management Information Systems. 12, 4, 5–33 (1996).
179. Wilson, E.O.: The meaning of human existence. WW Norton & Company, New York NY (2014).
180. Wood, A.W.: Kantian ethics. Cambridge University Press Cambridge (2008).
181. Yoon, J. et al.: Bayesian model-agnostic meta-learning. Advances in neural information processing systems. 31, (2018).
182. Zäschke, T. et al.: Optimising conceptual data models through profiling in object databases. Presented at the International Conference on Conceptual Modeling (2013).
183. Zhang, C., Lesser, V.: Coordinating multi-agent reinforcement learning with limited communication. Presented at the Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems (2013).
184. Zhang, R. et al.: Discovering data quality problems. Business & Information Systems Engineering. 61, 5, 575–593 (2019).
185. Zhang, Y. et al.: Siren's song in the AI ocean: a survey on hallucination in large language models. arXiv preprint arXiv:2309.01219. (2023).
186. Zhao, H. et al.: Explainability for large language models: A survey. ACM Transactions on Intelligent Systems and Technology. 15, 2, 1–38 (2024).
187. Zhu, T., Yang, Y.: Research on immersive interaction design based on visual and tactile feature analysis of visually impaired children. Heliyon. 10, 1, (2024).
188. Živković, S., Karagiannis, D.: Towards metamodelling-in-the-large: Interface-based composition for modular metamodel development. In: Enterprise, Business-Process and Information Systems Modeling. pp. 413–428 Springer (2015).