Предрегистрация методики НИОКР в блокчейне и семантический поиск по СКВ-матрицам: опыт платформы Academic Gate

2 нояб. 2026 г. · @Dr. Sergey Kibalnikov

С. В. Кибальников, Р. Г. Усманов

Международная научно-исследовательская лаборатория стратегии управления социально-технологических систем, МНИИПУ

УДК 001.8:004.75:025.4.03

Аннотация

В статье предложен протокол, в котором методика опытно-конструкторского или научно-исследовательского эксперимента (НИОКР) фиксируется в блокчейне до его начала, а результаты публикаций индексируются через СКВ-матрицы (SKW: «слово — что это — потери конкурента — ноу-хау»). Протокол реализован на платформе Academic Gate (academicgate.su), использующей блокчейн Erachain: гипотеза, состав команды, инструменты, план анализа и пороговые критерии успеха получают SHA-256-отпечаток, подпись автора, номер блока и время до сбора первых данных. Тем самым задним числом нельзя изменить ни гипотезу, ни критерии успеха, а вклад каждого участника («пай») закреплён до патентования. Вторая часть работы посвящена научным электронным библиотекам, где значительная доля полных текстов доступна лишь как изображения страниц и поиск ограничен аннотациями и ключевыми словами. Предложен конвейер «OCR → извлечение понятий → генерация СКВ-матрицы → экспертная верификация → хеширование → индексирование», переводящий поиск со слов на связки «проблема → решение». Приведены три СКВ-матрицы (платформы, протокола предрегистрации и библиотечного поиска), семидольный план пилота с МФТИ и измеримые критерии: рост проводимости $P \geq 20$ %, снижение потерь G на 15 %, ошибка прогноза потенциала $H \leq 12$ %. Обсуждаются ограничения: хеш доказывает существование и неизменность документа, но не истинность его содержания и не заменяет патент.

Ключевые слова: предрегистрация исследований, НИОКР, блокчейн, Erachain, SHA-256, фиксация авторства, воспроизводимость, СКВ-матрица, семидольник, семантический поиск, оптическое распознавание текста, научные электронные библиотеки.

Abstract

The paper proposes a protocol in which the design of an R&D experiment is anchored on a blockchain before the experiment begins, while publications are indexed through SKW matrices (keyword — definition — competitor losses — know-how). The protocol is implemented on the Academic Gate platform (Erachain): hypothesis, team, instruments, analysis plan and success thresholds receive a SHA-256 fingerprint, the author’s signature, a block number and a timestamp before data collection. For digital libraries whose full texts are often available only as page images, an OCR-to-SKW pipeline is proposed that shifts search from words to problem-to-

solution links. Three SKW matrices, a seven-stage pilot plan and measurable success criteria are presented, together with the limits of the approach.

Keywords: preregistration, R&D, blockchain, Erachain, SHA-256, authorship timestamping, reproducibility, SKW matrix, semantic search, OCR, digital libraries.

1. Введение

Междисциплинарные коллективы теряют результат в трёх местах: при установлении приоритета, при проверке выводов и при поиске чужих решений. Каждая из этих потерь имеет известную природу, но обычно устраняется разными инструментами, не связанными между собой.

Потеря приоритета. Споры о первенстве сопровождают науку с момента её институционализации; Р. Мертон показал, что они вытекают из самой системы признания, а не из личных качеств учёных [11]. На стадии НИОКР, когда патентной заявки ещё нет, идея, высказанная в докладе или переписке, практически не защищена. Криптографическая отметка времени, предложенная С. Хабером и У. Сторнеттой [8] и реализованная в распределённом реестре [12], позволяет доказать, что документ существовал в данном виде в данный момент, без доверия к третьей стороне.

Потеря воспроизводимости. Массовая проверка психологических исследований подтвердила лишь около трети исходных эффектов [15]; одна из причин — гибкость анализа после получения данных и подгонка гипотез под результат [9]. Ответом стала предрегистрация: гипотеза и план анализа публикуются до сбора данных [14]. Однако реестры предрегистрации централизованы и почти не применяются к прикладным НИОКР в вузах и на предприятиях.

Потеря найденного. Поиск в научных электронных библиотеках опирается на метаданные, аннотации и полнотекстовый слой. Если полный текст представлен изображениями страниц, содержательный поиск сводится к аннотации и авторским ключевым словам. Даже при наличии текста лексический поиск возвращает документы со словами запроса, а не с решениями задачи читателя.

Цель работы — описать единый протокол, в котором одна и та же запись в блокчейне закрепляет приоритет, делает эксперимент проверяемым и служит индексом для поиска по связкам «проблема → решение». Задачи: (1) формализовать перечень сведений, фиксируемых до эксперимента; (2) определить СКВ-матрицу как единицу индексирования; (3) предложить конвейер её автоматического построения для текстов-изображений; (4) задать измеримые критерии для пилота.

2. Материалы и методы

2.1. Платформа Academic Gate

Academic Gate (academicgate.su) — среда исследователя, собирающая через API сервисы дружественных сайтов вокруг общего блокчейна Erachain, работающего с 2020 г. [4]. Ранее содержимое реестра было доступно через интерфейс просмотра Academic Lab; Academic Gate добавила к нему модули команд и проектов, публикаций, анонимных опросов, ИИ-помощника (GigaChat) и каталог дистанционных курсов. Вход единый для обоих сайтов и связан с ключом автора в блокчейне. По оценке разработчиков, затраты

вычислительных ресурсов на поддержание записей в Academic Gate примерно в 100 раз ниже, чем в Academic Lab, при общем реестре.

Единица фиксации — запись R, связывающая документ D с автором и временем:

```
R = \langle \mathrm{SHA\text{-}256}(D),\; \sigma_{\text{автор}},\;
n_{\text{блок}},\; t \rangle
```

где σ — подпись ключом автора, n — номер блока, t — время. Хеш-функция SHA-256 [13] даёт 256-битный отпечаток, по которому нельзя восстановить текст, но любое изменение документа меняет отпечаток. Поэтому сам текст может храниться вне цепи (off-chain), а в цепь попадает только отпечаток.

2.2. Протокол предрегистрации НИОКР

Ключевое требование протокола: всё, что определяет интерпретацию результата, фиксируется до получения первых данных. Состав предрегистрации приведён в табл. 1.

Таблица 1. Сведения, фиксируемые в блокчейне до эксперимента

Элемент	Что фиксируется	Кто подписывает	Момент фиксации
Цель и гипотеза	Текст проекта по семидольнику, проверяемые гипотезы H1…Hk	Все участники	До набора выборки
Команда и паи	Состав группы, роли, вклад каждого («пай»: методика, ПО, данные, площадка)	Каждый участник за свой пай	До набора выборки
Методика сбора данных	Протоколы срезов К1–К4, схема датасета 3 × 3, анкеты	Автор методики	До первого среза
Инструментарий	Перечень ПО и оборудования, версии, параметры калибровки	Ответственный за контур	До первого среза
План анализа	Формулы метрик P, G, H, статистические критерии, правила исключения данных	Автор и научный руководитель	До первого среза
Критерии успеха	Пороговые значения: P ≥ 20 %, G ↓ 15 %, H ≤ 12 %	Руководитель проекта	До первого среза
Данные срезов	Отпечаток каждой волны данных под анонимными кодами	Оператор сбора	Сразу после волны
Отчёт	Хеш рукописи, отклонения от плана с обоснованием	Авторы	После анализа

Порядок работы с записью:

1. Автор загружает документ; платформа вычисляет SHA-256.
2. Автор подписывает отпечаток ключом Erachain; запись получает номер блока и время.
3. Любое изменение плана оформляется новой записью со ссылкой на предыдущую; прежняя версия остаётся видимой.
4. После эксперимента отчёт сопоставляется с предрегистрацией; расхождения выносятся в отдельный раздел «Отклонения от протокола».

Так устраняется возможность подогнать гипотезу или пороги под полученный результат: дата плана в реестре предшествует дате первой волны данных.

2.3. Модель потока и метрики P, G, H

Деятельность участника описывается балансом

```
N = P + G
```

где N — полный поток усилий, P — проводимость (доля, превращённая в результат), G — потери (ошибки, переключения внимания, прокрастинация). Когнитивный потенциал H — способность участника выдать результат; метрика качества модели — ошибка прогноза H относительно факта.

Проводимость оценивается как коэффициент регрессии отклика Y на воздействие Z без свободного члена:

```
P = \frac{\sum_{i} Y_i Z_i}{\sum_{i} Z_i^{2}}
```

Эта оценка предложена одним из авторов в докторской диссертации (1997) для статистической идентификации параметров единой теории потоков (ЕТП) — модели движения потоков энергии и вещества между фиксированными точками пространства [3]. В исходной постановке Y — изменение управляемого параметра на i-м слое пространства за шаг времени, Z — функция значений параметра на предыдущем шаге в том же и соседних слоях (i − 1, i, i + 1), P — коэффициент линейной регрессии. Именно поэтому нужны три точки по пространству: центральная и две соседние.

Так по уровням воды в трёх смежных створах определялся коэффициент шероховатости канала, а по температурам трёх смежных слоёв — коэффициент температуропроводности почвы. При верификации на рисовых оросительных системах Краснодарского края относительная погрешность расчёта уровней воды составила 2–3 %, средняя квадратическая ошибка расчёта температуры — 0,3–0,6 °C [3]. Метод уже тогда периодически адаптировал модель к объекту внутри контура управления; в настоящей работе он перенесён на социотехнические системы, а такт времени задаётся блоками реестра.

Эта оценка не требует исходной обучающей выборки: управление начинается с начальной оценки P_0, а модель обучается в процессе эксплуатации. Накопленные суммы обновляются рекуррентно после каждого шага по времени k:

```
S_{YZ}(k) = S_{YZ}(k-1) + \sum_{i=1}^{3} Y_{ik} Z_{ik}, \qquad S_{ZZ}(k) =
S_{ZZ}(k-1) + \sum_{i=1}^{3} Z_{ik}^{2}, \qquad P_k =
\frac{S_{YZ}(k)}{S_{ZZ}(k)}
```

где i — номер пространственной точки измерения. Начальные значения $S_{YZ}(0) = \varepsilon \cdot P_0$ и $S_{ZZ}(0) = \varepsilon$ задают вес априорной оценки; при $\varepsilon = 0$ рекуррентная оценка совпадает с $P = \Sigma YZ / \Sigma Z^2$. По текущей оценке P_k формируется следующее управляющее воздействие Z, а для отслеживания дрейфа суммы перед обновлением можно умножать на коэффициент забывания $\lambda \leq 1$.

Для работы рекурсии необходимы три измерения по времени и три измерения параметра в пространстве, то есть датасет 3 × 3. Каждый шаг по времени совпадает с блоком реестра Erachain, а блоки формируются раз в 30–60 с. Поэтому три шага по времени — это три

последовательных блока, и первая оценка по данным появляется через 1,5–3 мин после начала эксплуатации. Отметка блока одновременно задаёт момент измерения и подтверждает его неизменность, а хеш каждой итерации оценки, записанный в реестр, делает историю обучения модели проверяемой. Четыре среза К1–К4 соответствуют квадрантам методики СКВ и фиксируют режим деятельности участника в момент измерения.

2.4. СКВ-матрица как единица индексирования

СКВ-матрица (SKW) — таблица, каждая строка которой описывает одно ключевое понятие работы четырьмя полями с чередованием типов П–Л–П–Л:

1. **Ключевое слово (П)** — понятие, вокруг которого строится результат.
2. **Что это? (Л)** — рабочее определение в терминах статьи.
3. **Потери конкурента G (П)** — чего не учитывают существующие подходы.
4. **Ноу-хау G → P (Л)** — как предложенное решение превращает эти потери в проводимость.

Строка матрицы — минимальная связка «проблема → решение». Именно её, а не отдельное слово, предлагается индексировать и искать. План проекта задаётся семидольной матрицей: семь этапов (состояние вопроса, теория, методы и способы, проектирование, конструирование, внедрение, оценка результатов) по пяти вопросам ЗККРК — зачем, как, кто, когда, какими ресурсами.

Метод родствен приёмам ТРИЗ, где изобретательская задача формулируется через противоречие существующей системы [1], но отличается обязательной ссылкой на конкретные потери аналогов и фиксацией матрицы в реестре.

2.5. Конвейер автоматического построения СКВ-матриц

Для библиотек с полными текстами в виде изображений предлагается конвейер из шести шагов:

1. **Распознавание.** OCR страниц (например, Tesseract [16]) с сохранением координат строк; формулы и таблицы выделяются в отдельные объекты.
2. **Разметка структуры.** Определение аннотации, введения, методики, выводов; именно они дают поля «потери» и «ноу-хау».
3. **Извлечение понятий.** Языковая модель (класса BERT [7] или GigaChat) выделяет кандидатов в ключевые слова и их определения с указанием страницы-источника.
4. **Генерация строк.** Для каждого понятия модель в режиме поиска с дополнением [10] формулирует «потери конкурента» и «ноу-хау» только по фрагментам самой статьи, с цитатной привязкой к ним.
5. **Верификация.** Автор или эксперт подтверждает либо правит строки; неподтверждённые строки помечаются как машинные.
6. **Фиксация и индекс.** Подтверждённая матрица хешируется и записывается в реестр; строки индексируются как документы с полями (k, d, g, h) и векторными представлениями для смыслового поиска.

Матрицы хранятся в машиночитаемом виде, что соответствует принципам FAIR — находимость, доступность, совместимость и повторное использование [17].

3. Результаты

3.1. Записи первого года работы

За первый год в реестре зафиксированы работы разных жанров — от курсовой до монографии и внешней публикации; примеры приведены в табл. 2 [4]. Каждая запись содержит отпечаток SHA-256, подпись автора, номер блока и время и проверяется в обозревателе реестра на academicgate.su.

Таблица 2. Примеры записей Erachain на платформе Academic Gate

Номер записи	Что закреплено	Особенность
3413874-1	Статья Д. Ермолаева «ИИ-агент победил LaTeX»	Внешний автор, публикация с согласия; текст и отпечатки иллюстраций
3412688-1	Рукопись монографии Р. Г. Усманова «Биткоин как социально-технологическая система» [5]	Полный семидольник проекта принят научным руководителем
3412643-1	Курсовая работа «Искусственный интеллект в палеографии»	Учебная работа студента
3403966-1	Раздел «Контроль и оценка» проекта Academic Gate	Раздел принят руководителем по методике платформы
3360914-1	Проект «Лаборатория креативного ИИ»	Описание исследовательского направления

3.2. Предрегистрированный план пилота

На основе протокола подготовлен план пилота с МФТИ на 10–20 студентах продолжительностью 9 недель (табл. 3). План фиксируется в реестре целиком до этапа «Внедрение».

Таблица 3. Семидольная матрица пилота × ЗККР

Этап	Зачем	Как	Кто (пай)	Когда	Ресурсы
1. Состояние вопроса	Зафиксировать разрозненность подходов и потери G	Аудит паёв, стенограммы, интервью	С. В. Кибальников	Неделя 1	Время команды, записи
2. Теория	Единый язык N = P + G	Семидольник, СКВ-глоссарий	Т. А. Бородулина, методологи	Недели 1–2	Публикации, диссертация
3. Методы и способы	Превратить G в P	Цикл СКВ: слово → потери аналога → ноу-хау	Е. Н. Коржинская, И. В.	Недели 2–3	«Физиолаб», ритм-бот, симулятор «Шмель»
4. Проектирование	Спроектировать пилот	Группы, метрики, предрегистрация	Михаил Алексеевич,	Недели 3–4	Academic Gate, студенты

			С. В. Кибальников		
5. Конструирование	Собрать измерительный контур	Ритм-бот + ЭЭГ + «Шмель» + сертификаты авторства	Все паи	Недели 4–5	Оборудование, ПО, согласия участников
6. Внедрение	Запустить пилот	4 недели сессий, срезы до и после	Координатор (уточняется)	Недели 5–8	Учебная площадка
7. Оценка результатов	Проверить пороги P ≥ 20 %, G ↓ 15 %, H ≤ 12 %	Отчёт, сверка с предрегистрацией, NPS	С. В. Кибальников, Е. Н. Коржинская	Неделя 9	Аналитика, публикация

Критерии успеха и способы их измерения (табл. 4) входят в предрегистрацию и после фиксации не меняются.

Таблица 4. Критерии успеха пилота

Метрика	Порог	Как измеряется	Источник данных
Проводимость P	рост ≥ 20 %	Регрессия $\Sigma YZ/\Sigma Z^2$, до и после	Срезы СКВ, физиология
Потери G	снижение ≥ 15 %	Время возврата к задаче, число ошибок	Логи, наблюдение
Ошибка прогноза H	≤ 12 %	Прогноз против факта	Модель H
Фиксация артефактов	100 %	Сверка хешей	Реестр Academic Gate
Лояльность участников (NPS)	порог уточняется	Анкета после пилота	Анонимный опрос

3.3. СКВ-матрицы

Ниже приведены три матрицы: самой платформы, протокола предрегистрации и библиотечного поиска. Каждая строка — индексируемая связка «проблема → решение».

Матрица 1. Платформа Academic Gate и модель потока

№	Ключевое слово (П)	Что это? (Л)	Потери конкурента G (П)	Ноу-хау G → P (Л)
1	Когнитивный потенциал H	Способность участника выдать результат	Тесты интеллекта дают статичный срез и теряют динамику	Срезы СКВ К1–К4 с калибровкой по физиологии (ЭЭГ, вариабельность ритма сердца)
2	Проводимость P	Доля усилий, превращённая в результат: «стимул → результат»	Время реакции измеряют изолированно, вне контекста задачи	Регрессия $\Sigma YZ/\Sigma Z^2$ на датасете 3 × 3 с биологической обратной связью

3	Потери G	Ошибки, переключения внимания, прокрастинация	Остаточное внимание после переключения не учитывается	СКВ-глоссарий потерь и фиксация каждой волны данных в реестре
4	Пай	Вклад участника: методика, ПО, данные, площадка	Авторство теряется до подачи патентной заявки	Подписанный хеш вклада и сертификат авторства
5	Ритм, тайминг	Базовая функция мозга, миллисекундный диапазон	Тренировка ритма ведётся «на слух», без объективной метрики	Окна попадания с подкреплением; блоки реестра как такт времени

Матрица 2. Предрегистрация НИОКР в блокчейне

№	Ключевое слово (П)	Что это? (Л)	Потери конкурента G (П)	Ноу-хау G → P (Л)
1	Предрегистрация	Публикация гипотезы и плана анализа до сбора данных	Централизованные реестры охватывают в основном клинические и психологические исследования, но не прикладные НИОКР	Запись плана в общий реестр Erachain доступна любому коллективу
2	Гипотеза	Проверяемое утверждение с направлением эффекта	Гипотеза формулируется после результатов (HARKing)	Дата гипотезы в блоке предшествует дате первой волны данных
3	Критерий успеха	Числовой порог, при котором результат признаётся	Пороги выбираются задним числом	Пороги P, G, H входят в подписанный план
4	Отклонение от протокола	Изменение плана по ходу работы	Изменения не видны рецензенту	Каждая правка — новая запись со ссылкой на прежнюю
5	Приоритет	Доказанное первенство идеи	До патента идея, высказанная в докладе, не защищена	Отметка времени подтверждает существование текста на дату; патент не заменяется, но приоритет не теряется

Матрица 3. Поиск в научной электронной библиотеке

№	Ключевое слово (П)	Что это? (Л)	Потери конкурента G (П)	Ноу-хау G → P (Л)
1	Текст-изображение	Полный текст, доступный только как скан страниц	Поиск ограничен аннотацией и ключевыми словами автора	OCR с сохранением координат и разметкой разделов

2	Лексический поиск	Сопоставление слов запроса и документа	Выдаются статьи со словами запроса, а не с решениями задачи	Индекс по строкам СКВ (k, d, g, h) и их векторным представлениям
3	Машинная матрица	СКВ-матрица, построенная языковой моделью	Модель может приписать статье несуществующее решение	Генерация только по фрагментам статьи с привязкой к странице; подтверждение автором
4	Подлинность текста	Совпадение читаемого текста с опубликованным	Читатель не может проверить, менялась ли статья	Хеш текста и матрицы в реестре; кнопка проверки в карточке
5	Сравнение решений	Сопоставление подходов разных авторов к одной проблеме	Читатель вручную читает десятки аннотаций	Группировка выдачи по полю «потери конкурента»

3.4. Пример поиска по СКВ-матрицам

Запрос читателя: «как повысить вовлечённость студентов в онлайн-обучении, если расписание не учитывает индивидуальные ритмы внимания». Система разбирает запрос на проблему (поле g: «расписание не учитывает ритмы») и цель (поле k: «вовлечённость»), затем ищет строки, где проблема совпадает по смыслу, а ноу-хау её устраняет. В выдаче читатель видит не список статей, а строки вида «вовлечённость → групповые расписания игнорируют циркадные ритмы → персонализация расписания по пикам продуктивности», со ссылкой на страницу статьи и статусом записи в реестре. Лексический поиск по тому же запросу вернул бы все статьи со словами «вовлечённость» и «онлайн», включая не содержащие решения.

4. Обсуждение

Протокол меняет порядок доказательства: проверяется не только итог, но и то, что было обещано до эксперимента. Вместе с тем у подхода есть границы, которые следует оговаривать в каждой публикации.

Хеш доказывает существование, а не истинность. Совпадение отпечатков подтверждает, что текст не менялся с момента записи и существовал на дату блока [8]. Оно ничего не говорит о корректности данных и выводов. Предрегистрация снижает свободу подгонки, но не заменяет рецензирование и повторение эксперимента.

Запись не заменяет патент. Отметка времени закрепляет авторство и дату, но не создаёт исключительного права. Её роль — не потерять автора на пути к патенту или ноу-хау.

Персональные данные остаются вне цепи. Физиологические и поведенческие показатели участников относятся к персональным данным [6]; в реестр попадают только отпечатки обезличенных наборов, а сами данные хранятся в изолированных контурах с шифрованием. Необратимость записи исключает удаление по требованию субъекта, поэтому записывать в цепь сами данные недопустимо.

Ошибки распознавания и генерации. OCR даёт ошибки на формулах, таблицах и старых сканах; языковая модель может приписать статье решение, которого в ней нет. Поэтому машинные строки матрицы маркируются отдельно, каждая привязана к странице

источника, а в индекс с полным весом попадают только строки, подтверждённые автором или экспертом.

Риск смещения выдачи. Если в ранжировании поощрять работы с записью в реестре, это стимулирует фиксацию, но может вытеснить из выдачи старые и внешние работы, не прошедшие её. Статус записи разумно показывать как признак проверяемости, а не как меру качества, и оставлять сортировку по релевантности основной.

Метрики пилота пока плановые. Пороги P, G и H — гипотезы, зафиксированные заранее; их подтверждение или опровержение и станет первым содержательным результатом протокола.

5. Заключение

1. Предложен протокол предрегистрации НИОКР, при котором гипотеза, команда с паями, методика, инструментарий, план анализа и пороговые критерии подписываются и фиксируются в блокчейне Erachain до первого среза данных.
2. Одна запись решает три задачи: закрепляет приоритет до патента, делает результат проверяемым по исходному плану и служит единицей индекса.
3. СКВ-матрица определена как набор строк «ключевое слово — определение — потери аналогов — ноу-хау»; поиск по строкам переводит выдачу со слов на связки «проблема → решение».
4. Для библиотек с текстами-изображениями предложен конвейер OCR → разметка → извлечение → генерация → верификация → хеширование и индекс, с обязательной привязкой машинных строк к странице источника.
5. Подготовлен предрегистрированный план пилота с МФТИ (9 недель, 10–20 участников) с порогами $P \geq 20$ %, снижение G на 15 %, $H \leq 12$ %; его проведение станет первой проверкой протокола.

Дальнейшая работа: проведение пилота, оценка точности автоматических СКВ-матриц против экспертных и пилотное подключение протокола к открытой научной электронной библиотеке.

Список литературы

1. Альтшуллер Г. С. Найти идею: введение в теорию решения изобретательских задач. — Новосибирск: Наука, 1986. — 209 с.
2. ГОСТ Р 7.0.100–2018. Библиографическая запись. Библиографическое описание. Общие требования и правила составления. — М.: Стандартинформ, 2018.
3. Кибальников С. В. Совершенствование управления рисовыми оросительными системами: автореф. дис. … д-ра техн. наук: 06.01.02. — Бишкек: Киргизский сельскохозяйственный институт им. К. И. Скрябина, 1997.
4. Кибальников С. В., Усманов Р. Г. Academic Gate: платформа для междисциплинарных групповых исследований с фиксацией результатов в блокчейне [Электронный ресурс]. — 2026. — URL: https://academicgate.su
5. Усманов Р. Г. Биткоин как социально-технологическая система: рукопись монографии [Электронный ресурс] // Academic Gate. — Запись Erachain 3412688-1. — 2026.
6. Федеральный закон от 27.07.2006 № 152-ФЗ «О персональных данных».

7. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // Proceedings of NAACL-HLT 2019. — P. 4171–4186.
8. Haber S., Stornetta W. S. How to time-stamp a digital document // Journal of Cryptology. — 1991. — Vol. 3, No. 2. — P. 99–111.
9. Ioannidis J. P. A. Why most published research findings are false // PLoS Medicine. — 2005. — Vol. 2, No. 8. — e124.
10. Lewis P., Perez E., Piktus A. et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks // Advances in Neural Information Processing Systems 33 (NeurIPS 2020). — P. 9459–9474.
11. Merton R. K. Priorities in scientific discovery: a chapter in the sociology of science // American Sociological Review. — 1957. — Vol. 22, No. 6. — P. 635–659.
12. Nakamoto S. Bitcoin: A Peer-to-Peer Electronic Cash System [Электронный ресурс]. — 2008. — URL: https://bitcoin.org/bitcoin.pdf
13. National Institute of Standards and Technology. FIPS PUB 180-4. Secure Hash Standard (SHS). — Gaithersburg, 2015.
14. Nosek B. A., Ebersole C. R., DeHaven A. C., Mellor D. T. The preregistration revolution // Proceedings of the National Academy of Sciences. — 2018. — Vol. 115, No. 11. — P. 2600–2606.
15. Open Science Collaboration. Estimating the reproducibility of psychological science // Science. — 2015. — Vol. 349, No. 6251. — aac4716.
16. Smith R. An overview of the Tesseract OCR engine // Proceedings of the 9th International Conference on Document Analysis and Recognition (ICDAR 2007). — P. 629–633.
17. Wilkinson M. D., Dumontier M., Aalbersberg I. J. et al. The FAIR Guiding Principles for scientific data management and stewardship // Scientific Data. — 2016. — Vol. 3. — 160018.