

«Наименование учебного заведения»

«Факультет и кафедра»

«Название учебной дисциплины»

## ПРОЕКТ

На тему

«Персональный AI-менеджер по здоровью»

Выполнил:

ФИО и группа

Руководитель:

---

Город, год

# СОДЕРЖАНИЕ

## ВВЕДЕНИЕ

- 1 Теоретические основы персонального менеджмента здоровья и роль AI
- 2 Архитектура системы и функциональные модули ПАИМ
- 3 Источники данных, их качество и предобработка
- 4 Модели и алгоритмические подходы для управления здоровьем
- 5 Персонализация, профили пользователей и механизмы адаптации
- 6 Конфиденциальность, этика и регуляторные требования
- 7 Методы оценки эффективности, валидации и клинические и поведенческие исследования
- 8 Внедрение, экономическая модель и дорожная карта реализации

## ЗАКЛЮЧЕНИЕ

## СПИСОК ЛИТЕРАТУРЫ

## ВВЕДЕНИЕ

Актуальность темы персональных цифровых решений в здравоохранении растёт опережающими темпами ввиду демографических изменений, увеличения хронических заболеваний и доступности носимых сенсоров. Современные поколения устройств и платформ способны непрерывно собирать биометрические, поведенческие и клинические данные, однако их интеграция и трансформация в практически применимые рекомендации остаётся фрагментарной. В таких условиях персональный AI-менеджер по здоровью (далее — ПАИМ) представляет собой перспективную промежуточную систему между пациентом, медицинским персоналом и цифровой экосистемой: он объединяет сбор данных, анализ рисков, персонализированные рекомендации и сопровождение исполнения лечебных и профилактических программ. Это открывает перспективы для повышения приверженности лечению, раннего выявления обострений и оптимизации ресурсов здравоохранения, одновременно порождая технологические, этические и регуляторные вызовы.

Целью проекта является разработка концептуальной модели, архитектуры и критериев оценки эффективности ПАИМ, а также определение технологических и организационных мер, обеспечивающих достоверность, безопасность и соблюдение прав пациента. Для достижения цели мы ставим следующие задачи: 1) проанализировать теоретические основы и существующие решения в области цифровой медицины и AI-менеджмента здоровья; 2) предложить модульную архитектуру системы с учётом источников данных, алгоритмических компонентов и интерфейсов взаимодействия с пользователем и клиническими службами; 3) разработать подходы к персонализации рекомендаций и адаптивному обучению моделей на пользовательских данных; 4) сформулировать требования к приватности, безопасности и соответствию правовым нормам; 5) предложить методику клинической и технической валидации, а также план пилотного внедрения и оценку экономической целесообразности.

Структура проекта состоит из восьми глав, логика которых отражает путь от обоснования и концепции к технической реализации, оценке и внедрению. Первая глава рассматривает теоретические и прикладные предпосылки — понятия персонализированной медицины, менеджмента здоровья и роль AI. Вторая глава раскрывает архитектурную модель ПАИМ: функциональные блоки, каналы интеграции и требования к масштабируемости. Третья глава посвящена источникам данных и методам их предварительной обработки, объединения и семантической интерпретации. В четвёртой главе рассматриваются алгоритмические ядра: модели прогнозирования, обнаружения отклонений и механизмы объяснимости решений. Пятая глава фокусируется на персонализации: построении профилей пользователя, адаптивном машинном обучении и алгоритмах рекомендаций. Шестая глава анализирует правовые, этические и социальные аспекты внедрения, включая защиту персональных данных и проблемы алгоритмической справедливости. Седьмая глава описывает методику оценки эффективности, валидации и план клинических пилотных испытаний. Восьмая глава посвящена вопросам внедрения: интеграция с медицинскими организациями, бизнес-модель, экономическое обоснование и дорожная карта реализации.

Каждая глава содержит аналитический каркас, указывающий на ключевые предметы исследования, методы анализа и проблемные узлы, которые будут проработаны при дальнейшем развитии проекта. Предлагаемый подход сочетает доказательные медицинские принципы, современные практики машинного обучения и нормативно-правовые ограничения, что обеспечивает прикладную направленность и реализуемость предложенной архитектуры. Мы считаем, что системная проработка всех перечисленных аспектов позволит не только сформировать теоретическую основу для ПАИМ, но и подготовить практические рекомендации для его пилотного внедрения в российских условиях.

## **1 Теоретические основы персонального менеджмента здоровья и роль AI**

Глава раскрывает ключевые понятия: здоровье как многомерная категория, персонализированная медицина, менеджмент здоровья и цифровая трансформация медицинских служб. Мы рассматриваем здоровье не только как отсутствие болезни, но как динамическую систему, требующую постоянного мониторинга и адаптивных вмешательств, где персонализация базируется на биологических, поведенческих и социальных детерминантах. Цифровые технологии и методы искусственного интеллекта (глубокое обучение, машинное обучение) обеспечивают возможность обработки больших данных и выработки индивидуализированных рекомендаций, что показано в исследованиях по классификации кожных поражений и диагностике заболеваний с уровнем, сопоставимым с экспертами [8], а также в обзорах по deep learning и AI-методам [7; 6]. Важным выводом является то, что цифровой менеджмент здоровья потенцирует переход от реактивной к проактивной медицине, улучшая доступность диагностики и качество принятия решений, но одновременно порождает новые организационные и этические вызовы [5].

На основе указанной концепции мы формулируем функциональность персонально-адаптивной информационно-медицинской платформы (ПАИМ): непрерывный мониторинг состояния пациентов с применением сенсорики и анализом временных рядов, системы поддержки принятия клинических решений на основе алгоритмов прогнозирования рисков, цифровая терапия и сопровождение выполнения лечебных рекомендаций, а также профилирование и триаж для раннего выявления высоких рисков. Такая архитектура предполагает интеграцию многомодальных данных (электронные медицинские карты, геномика, данные wearables, социо-экономические показатели) и использование ансамблей моделей для повышения устойчивости выводов. Мы аргументируем выбор перечисленных функций логикой клинической полезности: мониторинг снижает временную задержку выявления ухудшения, CDS уменьшает

вариативность принятия решений, цифровая терапия повышает комплаенс и доступность интервенций [5; 12].

Аналитическая часть посвящена проблемам внедрения и эксплуатации AI-систем в здравоохранении, которые требуют системного решения. Во-первых, ограниченная прозрачность «чёрных ящиков» и неопределённость причинно-следственных выводов ставят под сомнение клиническую ответственность и пригодность для принятия автономных решений; это подтверждается критикой алгоритмических искажений и правовыми рисками [14; 10]. Во-вторых, вопрос надёжности и валидации моделей: многие разработки показывают высокую ретроспективную точность, но демонстрируют снижение эффективности в реальных условиях и на других популяциях, что требует проспективных исследований и рандомизированных испытаний в рамках регуляторных требований [2; 15]. В-третьих, риски утечки и неправильного использования персональных данных обуславливают необходимость внедрения технических гарантий (дифференциальная приватность, криптографические протоколы) и организационных мер, что описано в теоретических работах по приватности данных [13] и международных рекомендациях по этике AI в здравоохранении [11]. Мы считаем необходимым формализовать критерии отбора пилотных проектов: представительность тренировочных данных, прозрачность валидации, наличие процедур внешней аудиторской проверки и мониторинга постмаркетинговой производительности.

Краткая типология сценариев применения ПАИМ обосновывает приоритеты пилотирования. Во-первых, хронические неинфекционные заболевания (сердечно-сосудистые, сахарный диабет) показывают высокий потенциал экономии и клинической пользы благодаря мониторингу и прогностическим моделям; это подтверждается работами по машинному обучению в кардиологии [12]. Во-вторых, реабилитация и длительный уход выигрывают от адаптивных цифровых программ и дистанционного надзора, снижая госпитализации. В-третьих, превентивные и скрининговые программы с

использованием AI позволяют раннее выявление рисков в популяциях, однако требуют строгой калибровки для уменьшения ложноположительных результатов и социальных последствий. Наконец, профилирование пациентов для персонализации терапии — перспективная, но наиболее чувствительная к смещениям и требует последовательного наращивания доказательной базы. Исходя из анализа, мы предлагаем пилотировать проекты в сфере хронической патологии и реабилитации с поэтапной валидацией и применением технических и регуляторных мер по обеспечению безопасности, прозрачности и приватности

## **2 Архитектура системы и функциональные модули ПАИМ**

Глава предьявляет модульную архитектуру Персонального AI-менеджера по здоровью (ПАИМ), описывающую составные части системы, их назначение и взаимодействие. Мы рассматриваем ПАИМ как совокупность функциональных модулей: сбор данных, интеграция и хранение, аналитические ядра (диагностика, прогностические модели, персонализация рекомендаций), модуль интервенции (планирование и исполнение рекомендаций), интерфейсы взаимодействия с пользователем и медицинскими провайдерами, а также компоненты безопасности, управления доступом и аудита. Каждый модуль определён по входам, выходам и точкам интеграции: модуль сбора данных принимает входные потоки от носимых устройств, электронных медицинских карт, результатов лабораторий и анамнестических опросников; на выходе он нормализует данные и передаёт их в хранилище и шину событий для последующей обработки. Модуль интеграции обеспечивает преобразование форматов, сопоставление идентификаторов и семантическую привязку данных, гарантируя консистентность временных рядов и метаданных в соответствии с требованиями регуляторов [1,2]. Аналитические ядра принимают агрегированные данные и генерируют выводы: диагностические предположения, вероятностные прогнозы и персонализированные планы вмешательства. Для построения моделей мы опираемся на методы глубинного обучения и классические алгоритмы машинного обучения, применяемые в медицине [7,8,12], при этом реализуем механизмы версионирования моделей и валидации на отложенных выборках. Модуль интервенции трансформирует аналитические выводы в конкретные действия (напоминания, назначения к обследованиям, связи с медицинским персоналом) и отслеживает их исполнение через систему обратной связи от пользователя и устройств мониторинга.

Интеграция модулей организована по принципу событийно-ориентированной архитектуры с централизованным реестром сервисов: шина данных обеспечивает асинхронную доставку сообщений, а оркестратор отвечает

за маршрутизацию задач между анализатором, планировщиком и сервисами уведомлений. Важной частью архитектуры являются компоненты безопасности: аутентификация и авторизация с многофакторной проверкой, шифрование данных в покое и в канале, управление согласием пациента и журналы аудита с неизменяемой записью операций. Для защиты приватности мы предлагаем интегрировать методы дифференциальной приватности и приватного обучения при коллективной обработке данных [13], а также механизмы интерпретируемости и объяснимости моделей для уменьшения риска «чёрного ящика» и юридической непрозрачности [14]. Политики управления рисками и процедуры регулярного мониторинга моделей помогут выявлять и корректировать системные смещения, выявленные в эмпирических исследованиях [9,10]. Рамки этики и управления должны соответствовать международным рекомендациям ВОЗ и национальным нормативам [11,1,15].

Для практической реализации ПАИМ мы анализируем три варианта архитектуры. Первый — централизованная облачная платформа с мощными вычислениями для обучения и развертывания моделей: подходит для масштабируемой аналитики и быстрой интеграции данных, но требует строгого соблюдения требований по защите ПДн и сертификации медицинских систем [2,3]. Второй — гибридная архитектура с локальными агентами на устройстве пользователя, выполняющими персональные предобработки и обеспечивающими базовую приватность, при этом облако выполняет обучение и координацию; такой подход уменьшает поток чувствительных данных в облако и улучшает отзывчивость интерфейса. Третий — распределённая схема с федеративным обучением и приватными агрегаторами, при которой модели обучаются на местах, а централизованный сервис получает только агрегированные обновления; этот вариант усиливает приватность и соответствует рекомендациям по ответственному использованию больших данных в здравоохранении [13,11], но требует развитой инфраструктуры оркестрации и механизмов верификации качества моделей. Мы считаем, что

выбор варианта должен базироваться на анализе рисков, регулирования и требований к функциональности, сочетая технологическую эффективность с безопасностью и этичностью внедрения.

### **3 Источники данных, их качество и предобработка**

В рамках проекта по созданию персонального AI-менеджера по здоровью мы рассматриваем широкий набор источников данных: электронные медицинские записи (EHR), лабораторные результаты, данные с носимых и мобильных устройств, изображения и мультимодальные обследования, геномные и омные профили (при доступности), а также самоотчёты и социально-демографическая информация. Каждый тип источника предъявляет свои требования к сбору, валидации и предобработке: EHR и лабораторные данные обычно структурированы, но фрагментарны и зависят от локальных практик документирования; данные носимых устройств — плотные по времени, подвержены шуму и артефактам; геномные и омные наборы — высокоразмерные и чувствительные; тексты клинических записей требуют семантической нормализации и мэппинга к словарям медицинских понятий (например, SNOMED CT) и взаимодействию по стандартам обмена (HL7 FHIR). Для работы с медицинскими данными необходимо учитывать регуляторные и организационные рамки, сложившиеся в конкретной стране и в конкретных учреждениях; в РФ это регламентируется действующими федеральными актами и рекомендациями по цифровизации здравоохранения, что диктует требования к хранению, обмену и аудиту доступа к EHR [1–3,15].

Мы выстраиваем предобработку как многоэтапный конвейер: первичная валидация и нормализация, временная агрегация и выравнивание, обработка пропусков, удаление артефактов и приведение признаков к единому формату. Нормализация включает стандартизацию единиц измерения и приведение кодировок; временная агрегация требует выбора окон и методов (скользящие средние, экспоненциальное сглаживание) с учётом клинической интерпретируемости. Пропуски обрабатываются по характеру отсутствия: для случайно пропущенных значений используются множественная импутация и модели на основе ближайших соседей; для структурно отсутствующих показателей применимы индикаторные признаки и методы, сохраняющие

информацию о паттерне отсутствия. Для временных рядов носимых устройств целесообразны техники интерполяции и фильтрации, а также детекция и отбрасывание артефактов по правилам качества сигнала.

Оценка качества данных опирается на метрики полноты (completeness), корректности (correctness), своевременности (timeliness) и прослеживаемости происхождения (provenance). Мы определяем пороговые значения метрик для каждого источника и внедряем автоматические проверки качества на этапе ETL; при несоответствии данных требованиям применяется регламент возврата на источник или пометка для последующей ручной ревизии. Семантическая согласованность обеспечивается через мэппинг терминологии и валидацию кодов (например, контроль соответствия локальных кодов международным системам).

Подготовка признаков включает фильтрационные и обёрнутые методы отбора, регуляризацию и проверку стабильности признаков по кросс-средам; для объяснимости и борьбы с систематическими искажениями используется оценка важности признаков, калибровка моделей и тестирование на подвыборках, репрезентативных по ключевым демографическим критериям [6–8,12]. Особое внимание уделяется рискам предвзятости и дискриминации: выявление и корректировка перекосов в данных (например, недопредставленность групп) и аудит алгоритмов на предмет побочных эффектов, отмеченных в литературе

Конфиденциальность и защита чувствительной информации реализуются через шифрование, разграничение доступа, псевдонимизацию и современные методы приватности (дифференциальная приватность, secure computation) при обучении и публикации агрегированных результатов [13]. Выстраивая конвейер предобработки, мы стремимся обеспечить репродуцируемость, прозрачность и соответствие правовым требованиям, что является необходимым условием безопасного и этически корректного внедрения персонального AI-менеджера по здоровью [5,11].



#### **4 Модели и алгоритмические подходы для управления здоровьем**

В данной главе мы систематизируем набор алгоритмических решений, необходимых для реализации персонализированного AI-менеджера по здоровью (PAIM). Основные функциональные задачи включают прогнозирование рисков, детекцию трендов и обострений, выявление аномалий состояния пациентов, формирование рекомендаций и обеспечение интерпретируемости результатов для клиницистов и пациентов. Для каждой задачи мы рассматриваем пригодные классы алгоритмов, их достоинства и ограничения, а также требования к валидации и внедрению в клиническую практику.

Для прогнозирования рисков и классификации клинических исходов целесообразно использовать сочетание моделей классификации и регрессии. Традиционные методы (логистическая регрессия, деревья решений) обеспечивают прозрачность и простоту валидации, а сложные нелинейные модели (глубокие нейронные сети, градиентный бустинг) — более высокую предсказательную способность при больших объемах данных [6,7]. Важным практическим примером служит классификация поражений кожи с помощью свёрточных сетей, демонстрирующая возможности глубокого обучения в медицине при строгой клинической валидации [8]. Мы предлагаем применять многоуровневый подход: сначала — простые интерпретируемые модели для предварительного отбора признаков и объяснимых сигналов, затем — сложные модели для улучшения точности при обеспечении механизмов контроля и прозрачности результатов.

Задачи детекции трендов и прогнозирования обострений требуют методов анализа временных рядов. Здесь оправданы рекуррентные и трансформерные архитектуры для моделирования долгосрочных зависимостей, а также классические статистические методы (ARIMA, state-space модели) для краткой интерпретируемой прогностики. Комбинация статистических и нейросетевых подходов обеспечивает баланс между объяснимостью и адаптивностью модели к изменяющимся паттернам в данных.

Для обнаружения неожиданных состояний и сигналов тревоги необходимы алгоритмы аномалий: методы на основе плотности, модели автокодировщиков и методы обучения без учителя. В сценариях с редкими событиями мы рекомендуем использовать симуляции и генеративные модели для обогащения тренировочных выборок и повышения устойчивости детекции.

Повышение устойчивости прогнозов достигается с помощью методов ансамблирования (bagging, boosting, stacking), что снижает дисперсию и улучшает обобщающую способность моделей при гетерогенных клинических данных [6]. Для персонализации стратегий вмешательства целесообразно применять методы онлайн-обучения, байесовскую оптимизацию гиперпараметров и подходы с подкреплением для адаптивного подбора рекомендаций и режимов лечения в динамике пациента. Reinforcement learning может использоваться для оптимизации политик лечения при условии строгой симуляционной валидации и контролируемых испытаний [6,5].

Для задач рекомендательной логики мы предлагаем гибридные методы, объединяющие контентный и коллаборативный подходы, что позволяет учитывать и клинические характеристики пациента, и эмпирические паттерны популяции. Такой подход повышает релевантность рекомендаций и снижает риск переносимых предубеждений при использовании только популяционных моделей.

Особое внимание уделено объяснимости (XAI) и валидации: модели должны предоставлять интерпретируемые компоненты (вклад признаков, примеры-анalogии, локальные правила) и поддерживать требования клинической достоверности. Валидация алгоритмов должна соответствовать методологическим стандартам клинических исследований и цифровой медицины, включая оценку смещения и безопасности, что отражено в международных и национальных руководствах [4,5,11,15]. Важны также требования к защите данных и приватности, реализуемые через дифференциальную приватность и контролируемые процедуры доступа [13,14].

Таким образом, мы приходим к интегрированной архитектуре RAIM, где сочетание интерпретируемых моделей, глубоких нейросетей, методов временных рядов, аномалий и ансамблей, дополненное ХАІ и строгой валидацией, обеспечивает баланс между точностью прогноза, устойчивостью и клинической приемлемостью.

## **5 Персонализация, профили пользователей и механизмы адаптации**

В данной главе мы систематически рассматриваем методы построения персональных профилей здоровья и их использование для формирования адаптивных рекомендаций. В первую очередь описывается агрегирование долгосрочных и краткосрочных показателей: длительные биометрические тренды (например, хронизация показателей артериального давления или глюкозы) комбинируются с моментальными сигнатурами состояния (пульс, шаги, показатели сна) для получения многоуровневого представления о состоянии пользователя. В профиль также следует включать генетические маркеры и показатели среды (экологические, социоэкономические факторы), поскольку их учёт повышает предсказательную информативность моделей и позволяет объяснять межиндивидуальную вариативность ответов на вмешательства. Мы аргументируем комбинирование популяционных и индивидуально скорректированных моделей как стратегию, позволяющую использовать статистическую силу больших выборок при одновременной адаптации к особенностям конкретного пользователя: популяционные модели дают априорные распределения и шаблоны риска, а индивидуальные корректировки уточняют прогнозы на основе локальных данных пользователя, что снижает риск ошибочной калибровки при редких событиях. Теоретические и практические основания для применения глубоких нейронных архитектур и методов обучения для извлечения сложных биомедицинских признаков подтверждаются работами по глубокому обучению и его роле в медицине [7, 5, 8]. Для сегментации пользователей и кластеризации когорт мы описываем подходы, основанные как на классических алгоритмах (иерархическая кластеризация, k-means), так и на методах представления признаков через обучаемые эмбединги, что обеспечивает построение биометрически и поведенчески однородных когорт для последующей таргетированной интервенции. Контекстно-зависимые рекомендации формируются на основе текущего состояния, предсказанных траекторий и пользовательских

предпочтений, что позволяет формулировать интервенции с учётом риска и допустимости вмешательства для конкретного индивида. Оценка эффективности персонализации в этом блоке предусматривает количественные метрики (ROC/AUC для прогностических задач, прирост приверженности, снижение числа обострений и госпитализаций) и качественные индикаторы приемлемости рекомендаций для пользователя [12, 5]. Мы подчёркиваем необходимость прозрачности и интерпретируемости при переносе популяционных моделей на отдельных пользователей из-за возможных систематических сдвигов и ошибок обобщения [14, 9].

Во второй части главы мы анализируем механизмы адаптации моделей и архитектуры, обеспечивающие приватность и стабильную персонализацию. Федерированное обучение рассматривается как метод, позволяющий обучать глобальные модели без передачи сырых персональных данных на центральный сервер, что соответствует требованиям защиты медицинской информации и национальным регуляциям в области охраны здоровья [1, 2]; применение дифференциальной приватности при агрегировании обновлений рекомендуется для формального ограничения риска утечки сведений о конкретных пациентах [13]. Transfer learning предлагается как ускоряющий механизм персонализации: предварительное обучение на больших популяционных наборах обеспечивает быстрый старт для адаптации к отдельным пользователям при ограниченном локальном объёме данных, что уже показало эффективность в задачах медицинской классификации и прогнозирования [7, 8]. Reinforcement learning описывается как инструмент оптимизации стратегий сопровождения, способный учитывать динамическую обратную связь и индивидуальные отклики на интервенции; при этом мы отмечаем необходимость формулирования корректных функций вознаграждения и контроля за безопасностью политики в клиническом контексте [6, 5]. Для оценки эффективности адаптации мы предлагаем сочетание проспективных и ретроспективных метрик: увеличение приверженности, снижение числа обострений и ухудшений, улучшение исходов

по клиническим шкалам — как ключевые целевые индикаторы [12, 5]. Одновременно мы анализируем возможные источники систематической ошибки при переносе моделей — смещение выборки, историческая предвзятость данных и эффекты «чёрного ящика» алгоритмов — и на основании обзора литературы предлагаем меры валидации, мониторинга и управления риском (аудит моделей, регулярная перекалибровка, прозрачные отчёты по производительности) для снижения негативных последствий персонализации [9, 10, 14].

## **6 Конфиденциальность, этика и регуляторные требования**

В данной главе мы анализируем правовые и этические аспекты внедрения Персонального AI-менеджера по здоровью (ПАИМ) в российских и международных контекстах и формулируем требования к практикам сбора, обработки, хранения и доступа к персональным медицинским данным. Исходя из анализа действующего законодательства Российской Федерации и методических рекомендаций, обработка данных пациентов должна обеспечиваться в рамках федеральных законов и регламентов, включая требования о защите персональных данных и особенностях обработки сведений о здоровье [1–3,15]. На международном уровне ключевые принципы управления ИИ в здравоохранении — прозрачность, подотчетность, защита приватности и клиническая безопасность — отражены в руководящих документах ВОЗ и профессиональной литературе и служат ориентиром для локальной адаптации политик [11,5]. Мы считаем необходимым интегрировать оба уровня нормирования при разработке и внедрении ПАИМ.

Формулируя требования к процессам жизненного цикла данных, мы выделяем следующие обязательные элементы. Необходимо обеспечить правовую основу и информированное согласие для сбора и использования медицинских данных, применение принципа минимизации данных, разграничение прав доступа, регулярное логирование операций и аудит доступа. Технически это достигается обязательным шифрованием данных в покое и при передаче, применением механизмов аутентификации и разграничения ролей, а также использованием методов приватности по дизайну, включая дифференциальную приватность для агрегированных аналитических задач [13]. При наличии распределённых источников данных целесообразно применять подходы федеративного обучения для снижения объёма централизованного обмена сырыми данными, что уменьшает риски утечек и повышает соответствие принципу минимизации.

Этическая часть затрагивает вопросы алгоритмической прозрачности, справедливости и конфликта интересов. Алгоритмы ПАИМ должны сопровождаться описанием ограничений, верификацией на репрезентативных датасетах и процедурой объяснимости решений для клиницистов и пациентов; черные ящики в медицинских решениях недопустимы без надёжных механизмов контроля и интерпретируемости [14]. Мы указываем на риск систематических предубеждений при обучении моделей на исторических данных и необходимость аудита на предмет дискриминации по социальным, расовым или иным признакам, с учётом известных случаев негативного воздействия подобных алгоритмов в здравоохранении [9,10]. Ответственность за риски должна быть формализована: разработчик и медицинская организация обязаны иметь процедуры оценки и управления рисками, обязанности по уведомлению при инцидентах и механизмы человеческого контроля, включая возможность немедленного вмешательства и остановки автоматических действий системы.

Регуляторный блок включает конкретные требования к клинической валидации, регистрации и постмаркетинговому надзору. Мы считаем необходимым проведение этапной валидации: ретроспективная оценка точности и безопасности на независимых данных, проспективные клинические испытания для подтверждения клинической эффективности и безопасности, а также мониторинг в реальном времени после внедрения. Для продуктов, выполняющих функции медицинского изделия или влияющих на медицинские решения, обязательна регистрация и сертификация в соответствии с национальными процедурами и методическими рекомендациями по оценке рисков и качества [2,15]. В протоколах требуется предусмотреть процедуры сообщения о побочных эффектах и инцидентах, периодическую переоценку алгоритмов и план управления изменениями.

В заключение, мы предлагаем интегрированный подход: сочетание юридических требований, технических мер приватности и криптозащиты, прозрачности алгоритмов, обязательной клинической валидации и

регламентированного постмаркетингового контроля. Только такая многоуровневая система управления обеспечит соблюдение прав пациентов, снижение рисков и надёжную интеграцию ПАИМ в клиническую практику.

## **7 Методы оценки эффективности, валидации и клинические и поведенческие исследования**

В данной главе мы предлагаем комплексную методологию оценки эффективности персонального AI-менеджера (РАИМ), интегрирующую технические, клинические и поведенческие метрики, а также принципы постмаркетингового мониторинга и экономической оценки. Техническая валидация начинается с оценки стандартных характеристик алгоритма: точности, AUC-ROC, F1-меры, показателей precision/recall и калибровки прогнозов (calibration). Для диагностики и скрининга целесообразно включать сравнение с уровнями экспертов и публикациями по аналогичным задачам (например, результаты нейросетей в дерматологии как ориентир по валидации классификаторов) [8], а также анализ устойчивости к шуму и адаптивным изменениям входных данных. Важной составляющей является оценка надёжности и воспроизводимости: кросс-валидация на независимых ретроспективных когортах и внешняя валидация на данных других учреждений с детальной отчётностью по источникам данных и предобработке [7]. Техническая безопасность включает тесты на целостность модели, версионирование и логирование, а также оценку риска «чёрного ящика» и меры прозрачности (локальные и глобальные методы объяснения) в соответствии с требованиями прозрачности алгоритмов [14]. Для защиты персональных данных и минимизации риска утечки мы рекомендуем использование криптографических протоколов и при необходимости методов дифференциальной приватности при агрегации статистик [13]. Проблемы алгоритмической дискриминации и системных искажений требуют предварительного анализа смещения по демографическим и клиническим подгруппам, применения коррекционных методов и обязательного тестирования на справедливость, учитывая известные примеры предвзятости в системах здравоохранения [9,10]. Клиническая оценка должна опираться на проспективные исследования с клиническими конечными точками, а не только на улучшение промежуточных предикторов. Мы рекомендуем выбирать

основную конечную точку, измеряемую пациент- и клинико-ориентированными показателями: снижение частоты серьёзных осложнений, госпитализаций или времени до клинического события. Дизайн исследования может варьироваться от рандомизированного контролируемого испытания (индивидуального или кластерного при внедрении в клинику) до effectiveness–implementation дизайна для одновременной оценки эффективности и внедрения в реальной практике [15]. Методологически важны: расчёт размера выборки при  $\alpha=0,05$  и мощности 80–90 % на основе ожидаемого эффекта и базовой частоты события; учёт внутрикластерной корреляции при кластерных RCT; анализ по намерению лечить; применение смешанных линейных моделей и моделей пропорциональных рисков для временных исходов; сравнение ROC кривых методом ДеЛонга для диагностических задач. Обработка пропущенных данных — множественная импутация при предположении MAR с дополнительными сценариями чувствительности. Поведенческие и внедренческие метрики включают измерение приверженности рекомендациям, удобства использования, нагрузки на персонал, показателей внедрения (adoption, fidelity, reach, sustainability) и изменение клинической практики. Для оценки безопасности вводится Доска по мониторингу данных, протоколы быстрой реакции на побочные события и регулярные ревью производительности в условиях реального мира с детекцией дрейфа и периодической перекалибровкой модели. Экономическая оценка проводится через анализ затрат и результатов (ICER, бюджетное воздействие) с моделированием сценариев и анализом чувствительности, опираясь на выводы о потенциальной экономии и рисках автоматизации в здравоохранении [5]. Регуляторно-этические аспекты соответствуют национальным требованиям и рекомендациям по клинико-методическим исследованиям и внедрению цифровых технологий в здравоохранении, включая нормативы о защите прав пациентов и порядке клинических исследований в РФ [1,2,15], а также международные принципы этики и управления ИИ в здравоохранении, изложенные ВОЗ [11]. В сумме мы

предлагаем многоуровневый протокол валидации и оценки, сочетающий техническую строгость, клиническую релевантность и системный подход к внедрению и мониторингу РАИМ в реальной практике.

## **8 Внедрение, экономическая модель и дорожная карта реализации**

Глава описывает практические шаги внедрения персонального AI-менеджера по здоровью (ПАИМ) в систему здравоохранения и коммерческие сценарии, а также обосновывает экономическую модель и дорожную карту реализации. Мы предлагаем поэтапную стратегию, начиная с пилотных проектов в контролируемых условиях, переходя к масштабированию и устойчивой коммерциализации с учётом регуляторных, этических и технических факторов. Пилотные испытания целесообразно проводить в профильных медицинских организациях с участием мотивационных групп врачей и пациентов, что позволит собрать клинические данные для валидации алгоритмов, оценить интеграцию с локальными EHR и выявить операционные риски [3–4,12]. На этапе пилота критически важны протоколы согласия пациентов, механизмы анонимизации данных и технические решения по приватности, включая подходы дифференциальной приватности при обмене метаданными и обучении моделей [13,11]. Одновременно необходимо обеспечить соответствие национальному регулированию и стандартам, указанным в федеральных нормативных актах и методических рекомендациях по цифровизации здравоохранения [1–2,15].

Экономическая модель основывается на гибридном подходе финансирования. Мы предлагаем сочетание государственных субсидий для начального развертывания в государственных учреждениях, платных подписок и разовых лицензий для частных клиник, а также моделей оплаты за результат (outcome-based payment) для снижения финансового риска заказчиков. В структуру затрат входят разработка и валидация алгоритмов, интеграция с EHR, инфраструктура хранения данных, обеспечение кибербезопасности, обучение персонала и постпродажная поддержка. Оценка ROI должна учитывать как прямые экономии (снижение повторных госпитализаций, оптимизация расходных материалов, уменьшение рабочего времени врачей), так и косвенные эффекты (улучшение качества жизни пациентов, снижение заболеваемости), с горизонтом возврата инвестиций 3–5 лет в зависимости от масштаба внедрения

и исходных показателей клиники [5,12]. Для коммерческих партнёров целесообразно предусмотреть разделение доходов при использовании модели SaaS и интеграции платных сервисов телемониторинга.

Ключевые риски и меры их снижения включают алгоритмическую предвзятость, непрозрачность «чёрных ящиков», утечки данных и регуляторные ограничения. Необходимо внедрять процедуры постмаркетингового мониторинга алгоритмов, независимые аудиты и инструменты интерпретируемости, а также требования по тестированию на репрезентативных популяциях, чтобы минимизировать дискриминацию и искажения в клинических решениях [8–10,14]. Правовое и этическое сопровождение на всех этапах соответствует рекомендациям ВОЗ по этике и управлению ИИ в здравоохранении

Дорожная карта реализации включает несколько фаз: подготовительную (юридическое и техническое проектирование, подбор пилотных площадок), пилотную (полевые испытания, сбор и анализ данных, корректировки), фазу масштабирования (массовая интеграция в сети клиник, обучение кадров, стандартизация процессов) и фазу устойчивой эксплуатации (постоянная поддержка, обновление моделей, мониторинг показателей). Мы рекомендуем ввести KPI на каждом этапе: скорость внедрения, доля пользователей, точность диагностики, сокращение времени на приём, снижение повторных госпитализаций и экономия на пациент/год. Планы оценки воздействия должны включать сравнительные аналитические исследования до и после внедрения, экономические расчёты и отчёты по безопасности пациентов.

Для преодоления организационных барьеров необходима координация между министерствами здравоохранения, ИТ-провайдерами, медицинским сообществом и образовательными институтами для подготовки кадров и обновления клинических протоколов. Комплекс мер по подготовке персонала, прозрачные схемы финансирования и строгие требования к совместимости с EHR создадут основу для устойчивого внедрения, минимизируют риски

коммерциализации через ложные клинические процессы и обеспечат соответствие правовым и этическим стандартам [1–4,11,15].

## ЗАКЛЮЧЕНИЕ

В заключении мы обобщаем достижение поставленной цели и выполнение задач проекта по разработке концептуальной модели персонального AI-менеджера по здоровью. Цель проекта — создание целостного подхода к проектированию ПАИМ, сочетающего технологические, клинические и нормативные аспекты — достигнута через последовательную проработку теоретических основ, архитектуры, источников данных, алгоритмических решений, персонализации, этики и методики оценивания эффективности. По задачам: анализ существующей предметной области показал, что сочетание адаптивных AI-методов и доказательной медицины позволяет формировать прогнозы клинически значимой точности при условии качественных данных и соответствующей валидации; предложенная модульная архитектура обеспечивает гибкость выбора между облачными и гибридными решениями и позволяет учитывать требования приватности и локальную специфику инфраструктуры; рассмотренные методы сбора и предобработки обеспечивают основу для семантической интеграции разнотипных данных и повышения пригодности исходных сигналов для аналитики.

Сводные выводы по главам демонстрируют взаимную связанность проблем и решений: теоретические положения (глава 1) определяют функциональные требования и сценарии использования, которые легли в основу системной архитектуры (глава 2). Качество и разнообразие источников данных (глава 3) напрямую влияют на выбор алгоритмических подходов (глава 4) и на возможности персонализации (глава 5). Правовые и этические ограничения (глава 6) формируют границы допустимых технических решений и требуют встроенных механизмов приватности и объяснимости. Методики оценки (глава 7) дают инструментальные критерии для принятия решения о масштабировании, а практические аспекты внедрения и экономическая модель (глава 8) обеспечивают реалистичность перехода от прототипа к устойчивой услуге.

Практическое значение результатов состоит в формировании целостного проектного плана, пригодного для реализации пилотных проектов в российских условиях: модульная архитектура облегчает интеграцию с локальными EHR и адаптацию под требования здравоохранения РФ; предложенные процедуры валидации и мониторинга позволяют снизить риски неверных клинических решений и своевременно выявлять систематические отклонения. В научно-методическом плане проект вносит вклад в систематизацию требований к цифровым менеджерам здоровья и формирует набор критериев для оценки их клинической и экономической эффективности.

Рекомендации для дальнейшей реализации включают проведение пилотных РКИ в приоритетных клинических сценариях (например, сердечно-сосудистая медицина, сахарный диабет, постинсультная реабилитация), применение федеративных схем обучения для защиты приватности и развитие модулей объяснимости для клиницистов. Необходимо также выработать стандарты взаимодействия с регуляторами и смежными участниками экосистемы здравоохранения для ускорения процедур сертификации и возмещения затрат. Наконец, важно обеспечить механизм постоянного обновления моделей на основе реальных данных и обратной связи пользователей, что будет ключевым фактором долгосрочной эффективности и устойчивости ПАИМ.

## СПИСОК ЛИТЕРАТУРЫ

1. Федеральный закон от 21.11.2011 N 323-ФЗ «Об основах охраны здоровья граждан в Российской Федерации»
2. Приказ Министерства здравоохранения Российской Федерации (регламенты по обработке персональных данных в здравоохранении)
3. Гладков А.В. Искусственный интеллект в здравоохранении: современные подходы и практики. Москва: Медицинская книга, 2020
4. Смирнова Е.Н. Персонализированная медицина и цифровые технологии: методологические аспекты. Вестник клинической медицины, 2019, т. 12, №4
5. Topol E. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. New York: Basic Books, 2019
6. Russell S., Norvig P. Artificial Intelligence: A Modern Approach. 3rd ed. Pearson, 2010
7. LeCun Y., Bengio Y., Hinton G. Deep learning. Nature, 2015
8. Esteva A., Kuprel B., Novoa R.A. et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature, 2017
9. Obermeyer Z., Powers B., Vogeli C., Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science, 2019
10. O'Neil C. Weapons of Math Destruction. Crown Publishing, 2016
11. WHO. Ethics and governance of artificial intelligence for health. World Health Organization, 2021
12. Krittanawong C. et al. Machine learning in cardiovascular medicine: are we there yet? Journal of the American College of Cardiology, 2017
13. Dwork C., Roth A. The algorithmic foundations of differential privacy. Foundations and Trends in Theoretical Computer Science, 2014
14. Price W.N. II. Big Data and Black-Box Medical Algorithms. Journal of Law and the Biosciences, 2018
15. Руководящие документы и методические рекомендации по цифровизации здравоохранения в Российской Федерации (национальные программы и отраслевые стратегии)

Это пример работы выполненный нейросетью «Напишудзу», подробнее по ссылке: <https://reshudzu.ru/proekt>