HEALTH NET.
Инфраструктурный центр

ПЛАТФОРМА НТИ

академпарк

СЕЧЕНОВСКИЙ УНИВЕРСИТЕТ
НАУК О ЖИЗНИ

AIRI

ИССЛЕДОВАНИЕ ИСПОЛЬЗОВАНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА (ИИ) В ФАРМАКОЛОГИИ И DRUG DISCOVERY

Отраслевой аналитический обзор

2021

Научно-технологический парк Новосибирского Академгородка
ISBN 978-5-6056181-4-0

HEALTH NET.
Инфраструктурный центр

ПЛАТФОРМА НТИ

УДК 004.8:615:338.45
ББК 65.305.735:32.813
О-86

Редакционная коллегия:
М.Р. Галямова, О.В. Валиева

Авторы:
М.Р. Галямова, руководитель Центра национальной технологической инициативы по направлению «Хелснет»
О.В. Валиева, канд. экон. наук, старший научный сотрудник ИЭОПП СО РАН
К.М. Кравцов, эксперт Центра национальной технологической инициативы по направлению «Хелснет»
К.В. Песков, руководитель Центра математического моделирования в разработке лекарств ПМГУ им. Сеченова
А.А. Кадурин, Институт искусственного интеллекта AIRI

Отраслевой аналитический отчет «Исследование использования искусственного интеллекта (ИИ) в фармакологии и Drug Discovery» // — Инфраструктурный центр НТИ «Хелснет», Фонд «Технопарк Академгородка», Новосибирск, 2026

Исследование представляет собой комплексный анализ рынка использования искусственного интеллекта в фармакологии и разработке лекарственных средств (Drug Discovery) и охватывает как глобальные тренды, так и специфику российского сектора. В работе рассматриваются ключевые технологические драйверы, динамика инвестиций, деятельность ведущих игроков и стартапов, а также правовые аспекты регулирования отрасли. Особое внимание уделено российскому рынку: уровню его зрелости, барьерам роста и практическим кейсам внедрения ИИ в период 2022–2025 гг. Завершает работу глубокий библиографический анализ научных публикаций и методологических подходов в области драгдизайна за последние пять лет, позволяющий выявить актуальные векторы развития технологий.

Отраслевой аналитический отчет «Исследование использования искусственного интеллекта (ИИ) в фармакологии и Drug Discovery» подготовлен Фондом «Научно-технологический парк Новосибирского Академгородка», участвующем в Национальной технологической инициативе при поддержке Фонда поддержки проектов Национальной технологической инициативы и Министерства промышленности и торговли Российской Федерации, в рамках реализации Программы по развитию отдельного направления Национальной технологической инициативы «Хелснет».

Дизайн Е.А. Александрович, Д.Ю. Ульянова

© ИЦ НТИ «Хелснет»,
Научно-технологический парк Новосибирского Академгородка, 2026
При перепечатке ссылка обязательна

ISBN 978-5-6056181-4-0

Содержание

ВСТУПИТЕЛЬНОЕ СЛОВО 4

1. АНАЛИЗ ТЕКУЩЕГО СОСТОЯНИЯ И ПЕРСПЕКТИВ РАЗВИТИЯ РЫНКА ИИ В ФАРМАКОЛОГИИ И DRUG DISCOVERY 8

1.1. Глобальные тренды, технологии и ключевые игроки 8

1.2. Сегменты и динамика роста мирового рынка 14

1.3. Драйверы и ограничения развития рынка 16

1.4. Правовое регулирование на различных рынках 19

1.5. Инвестиции, кейсы и стартапы на глобальных рынках 22

2. РОССИЙСКИЙ РЫНОК ИИ В ФАРМЕ И DRUG DISCOVERY 28

2.1. Структура рынка и ключевые участники в России 28

2.2. Уровень зрелости и текущие практики применения ИИ 28

2.3. Инвестиции, партнерства и кейсы на российском рынке 29

2.4. Барьеры роста и точки развития 32

2.5. Правовое регулирование в России 33

2.6. Первоочередные шаги по формированию национальных гайдлайнов 36

2.7. Практики применения ИИ в российских проектах 2022–2025 гг. 38

3. БИБЛИОГРАФИЧЕСКИЙ ПОИСК 41

3.1. Методология исследования 41

3.2. Приоритеты в научных исследованиях за последние пять лет 41

3.3. Тренды в публикациях за 2022–2026 гг. 43

3.4. Анализ трендов в методологических подходах в исследованиях использования моделей ИИ в драгдизайне 46

МНЕНИЯ ЭКСПЕРТОВ 51

ЗАКЛЮЧЕНИЕ 59

СПИСОК ИСТОЧНИКОВ 60

ПРИЛОЖЕНИЕ 1 64

ПРИЛОЖЕНИЕ 2 77

Вступительное слово

ИИнновации в фарме. Год 2026 A.D.

Технологии (в широком смысле понимания термина из тезауруса закона о технологической политике) в фармации оказывают существенное влияние на научные исследования, фармразработку, клинические исследования, сами фармтехнологии как отдельный раздел, промышленное производство, продвижение и продажи. Фактически весь жизненный цикл погружен в информационные технологии (ИТ), а теперь, видимо, подошло время искусственного интеллекта (ИИ). Все поле начало «прошиваться» подходами и инструментами ИИ. Мы наблюдаем за развитием более 50 специализированных ИИ-решений для всех этапов жизненного цикла разработки от команд из разных стран. Для каждого этапа уже складывается свой набор инструментов, например: BenevolentAI и Open Targets — для поиска биомишеней; Schrödinger, Recursion, Atomwise — для ранних доклинических исследований; SYNTHIA — для ретросинтеза; Medidata, Veeva, Flatiron — для клинических исследований; Palantir и SAP — для производства и постмаркетингового анализа. Генеративный ИИ занимает в этой картине заметное место — прежде всего в работе с текстом, в научной коллаборации разработчиков, в регуляторной документации.

Если ретроспективно оценивать развитие фарминноваций, то видна устойчивая закономерность: качественные скачки развития совпадают со сменой базовых технологий. В рамках одной технологии идет постепенное накопление улучшений; точки прорыва возникают при переходе к следующей стадии зрелости — как в промышленных, так и в ИТ. Это видно как в пределах классического промышленного синтеза (органическая химия, ферментация), так и в биотехнологиях: рекомбинантные белки, моноклональные антитела, мРНК-платформы, генная и клеточная терапия, которые с недавнего времени доминируют в актуальном мировом пайплайне. По данным анализа одобрений FDA, на биотехнологические компании приходится 47% новых одобрений против 38% у традиционных фармкомпаний. В классе биологических препаратов биотехи занимают 63% — фактически весь сегмент за ними. В малых молекулах их доля 39%, и граница между биотехсегментом и классической фармой размывается даже там, где исторически доминировала большая фарма. Мировой рынок инноваций при этом устроен как рынок стартапов и сделок: 65% новых одобренных препаратов крупная фарма получает извне — через M&A и лицензирование, и только 28% составляет ее внутренний R&D. Таким образом, технологическое «давление» делает изменения быстрее и быстрее, и это существенно важно при разговоре о пришедших в индустрию ИИ-подходах и решениях.

Параллельно и внутри биотехволны ИТ переформатировали методологию разработки. Автоматизация лабораторий, специализированное программное обеспечение (ПО), математическое моделирование — молекулярная динамика, физико-химические расчеты, докинг — перевели существенную часть R&D в режим in silico. Циклы сжались (выигрыш во времени стал коммерческим параметром), стоимость гипотез существенно упала. Решение при этом остается за исследователем, пока ИТ выполняют роль инструмента поддержки.

Новый этап, который мы наблюдаем последние пару лет, по природе отличается от предыдущих. ИИ-модель учится на данных, обобщает их за пределами явно заданных правил, и новая обученная модель становится производственным активом R&D. Здесь важно сделать терминологическую оговорку: применение ИИ в фарме не сводится к генеративным моделям. Генеративный ИИ — лишь одна из семей моделей. Точек

приложения значительно больше: предиктивные модели свойств и токсичности; графовые нейросети для поиска биомишеней; foundation-модели для работы с омиксными данными; reinforcement learning для генерации молекулярных кандидатов; компьютерное зрение для патоморфологии и другие.

При всем богатстве отдельных решений мировая практика закономерно тяготеет к платформенной конструкции. Эта конвергенция следует из природы технологии: ценность ИИ растет нелинейно с объемом и связностью данных на разных этапах жизненного цикла фармразработки, когда принципиальна сквозная преемственность. Практики мировых регуляторов предъявляют серьезные требования к воспроизводимости и валидации моделей. Все это прозрачно и отслеживаемо на платформенном уровне. Сквозные ИИ-модули по горизонтали, точечные инструменты для каждого этапа по вертикали (матрица), единая база данных — похоже, теперь это фундамент для современной разработки фарминноваций.

Отдельного внимания заслуживают RWD- и RWE-технологии. Real-world data — данные о здоровье пациентов и оказании медицинской помощи, собираемые вне клинических исследований: электронные медицинские карты, регистры, страховые claims-базы. Real-world evidence — клинические доказательства, получаемые из анализа этих данных. Развитие идет от вспомогательной роли в постмаркетинговой безопасности к самостоятельному источнику доказательств для регуляторных решений — вплоть до замещения части контрольных групп в клинических исследованиях. В декабре 2025 года произошел значимый сдвиг: FDA сняло обязательное требование предоставлять идентифицируемые данные на уровне отдельного пациента для RWE по медицинским изделиям, открыв дорогу к работе с большими деидентифицированными базами — claims и регистрами, а также сообщило, что аналогичное расширение готовит для лекарств и биологических препаратов. В горизонте 2026–2027 гг. методологический коридор для RWE в фарме может стать существенно шире. Для нас с вами важно, что национальные регистры, интеграция с ЕГИСЗ и гибридные дизайны клинических исследований формируют новый «несущий слой» платформенной архитектуры, которой, судя по всему, не избежать.

Отдельным существенным пунктом разговора является регуляторика применения моделей ИИ в фармразработке, но это тема другого обзора.

Таким образом, активное использование подходов и моделей ИИ в фарме не очень похоже на моду или хайп. Управление временем, большими массивами данных, безопасность, точность и прослеживаемость — это задачи в первую очередь архитектуры.

Мы считаем, что основная и ближайшая наша задача состоит в том, как спроектировать и реализовать архитектуру так, чтобы Россия участвовала в наступающем технологическом «ИИ-сдвиге» как конкурентная страна-разработчик (не отдельные институты или фармкомпании) на собственной платформенной конструкции и создавала бестселлеры мирового уровня.

Олег Лавров
д. м. н., АНО «ИИЦ»
(экосистема компании «Иннопрактика»)

Вступительное слово

Когда мы говорим о внедрении ИИ в фармацевтическую разработку, важно сразу снять вопрос «зачем?». Для нас, как для компании, которая одновременно занимается и разработкой, и промышленным производством, ответ очевиден: чтобы быстрее, точнее и с меньшими рисками проходить путь от биологической гипотезы до лекарственного препарата, который дойдет до пациента. Это не про хайп. Это про экономику R&D, про управление сроками и про воспроизводимость результата.

В 2024–2026 гг. мы прошли путь от точечных пилотов к полноценным партнерским программам. Совместно со Сбером и Институтом AIRI мы запустили проекты по генеративному дизайну молекул для нейродегенеративных заболеваний, а также по созданию ИИ-решений, охватывающих поиск мишеней, доклиническую оценку и поддержку клинической разработки. Эти альянсы — не формальные соглашения. Это работающий контур, где модель учится на экспериментальных данных, а результат проверяется в лаборатории.

Что мы видим главным сейчас? Мировой рынок стремительно платформизируется. Разрозненные предикторы свойств или инструменты виртуального скрининга уступают место сквозной архитектуре, где интегрированы химический дизайн, биологические данные, расчет ADMET и даже элементы RWE-аналитики. Только такая конструкция позволяет использовать ИИ как производственный актив, а не как коллекцию «полезных игрушек».

Для России ключевой вопрос — не догнать, а встроиться в этот сдвиг на уровне собственной технологической базы. При этом мы прекрасно видим барьеры: дефицит стандартизированных данных, отсутствие нормативно закрепленных требований к валидации моделей и, что не менее важно, кадровый разрыв. Без специалиста, понимающего одновременно medicinal chemistry и машинное обучение (МО), любая платформа останется сырой.

Поэтому мы считаем своей задачей не только строить внутренние AI-компетенции, но и участвовать в создании отраслевых «правил игры». Именно нами совместно с партнерами из «Хелснет» НТИ, MS Decisions и AIRI разработано Рамочное соглашение о принципах использования ИИ в жизненном цикле лекарственных средств. Это попытка дать рынку ориентиры по качеству данных, валидации и управлению рисками — там, где государственное регулирование пока только формируется.

Конечно, ИИ не серебряная пуля. Модели ошибаются, данные дрейфуют, а final validation все равно остается за экспериментальными методами in vitro и in vivo. Но скорость, с которой меняется ландшафт, не оставляет выбора: либо ты становишься частью этой технологической волны, либо пропускаешь следующий цикл инноваций. Мы свой выбор сделали.

Владимир Андрианов
директор по работе с партнерами АО «Р-Фарм»,
заместитель лидера рабочей группы «Хелснет» НТИ

ЧАСТЬ I
АНАЛИЗ ТЕКУЩЕГО СОСТОЯНИЯ И ПЕРСПЕКТИВ РАЗВИТИЯ РЫНКА ИИ В ФАРМАКОЛОГИИ И DRUG DISCOVERY

1. Анализ текущего состояния и перспектив развития рынка ИИ в фармакологии и Drug Discovery

1.1. Глобальные тренды, технологии и ключевые игроки

В отчете рассматриваются глобальные тренды и ключевые игроки, разрабатывающие и внедряющие ИИ-продукты, применяемые в создании лекарственных препаратов на этапах discovery (этап поиска и первичного отбора лекарственных кандидатов), preclinical (доклинический этап) и clinical (клинические исследования). Фокус сделан на решениях, которые поддерживают ключевые R&D-процессы, выявление и приоритизацию терапевтических мишеней, дизайн и оптимизацию малых молекул и биологических препаратов, прогнозирование свойств и рисков на доклиническом этапе, а также анализ клинических данных для стратификации пациентов, выбора биомаркеров и повышения качества дизайна клинических исследований[1].

Для унификации терминов и обеспечения сопоставимости подходов используется общепринятая рамка жизненного цикла разработки лекарственных средств, в которой отмечается рост применения AI/ML-компонентов на некоммерческих этапах разработки и в клинических исследованиях, а также требования к контексту использования, управлению данными и практикам разработки моделей[2 3]. Дополнительно учитываются принципы, сформулированные европейским регулятором, где обсуждаются применение AI/ML на этапах разработки, требования к надежности, документации и оценке применимости в конкретных задачах жизненного цикла лекарственного препарата[4].

В 2024–2026 гг. глобальная повестка ИИ в разработке лекарств смещается от отдельных точечных моделей к платформенным подходам, которые можно масштабировать на разные задачи внутри цепочки поиска и разработки лекарств[5]. Центральным технологическим трендом стал рост количества и качества моделей-основ (foundation models), обучаемых на больших массивах химических, биологических и текстовых данных и далее дообучаемых под конкретные R&D-задачи, включая поиск мишеней, генерацию и оптимизацию молекул, прогнозирование свойств и рисков[6]. На практике это ведет к унификации ядра моделей и ускоряет перенос методов между терапевтическими областями, однако усиливает требования к качеству данных и валидации результатов на экспериментальных наборах.

При этом наблюдается быстрая эволюция данных моделей и основанных на них методов в сторону мультимодальности. Мировой тренд заключается в объединении нескольких типов данных в одной архитектуре и единой логике предобучения: химические структуры и реакции, последовательности и структуры белков, omics-профили, изображения (включая клеточные и гистологические), а также клинические признаки[7]. В обзорах 2025–2026 гг. мультимодальные предобученные модели описываются как новый базовый слой для задач drug discovery, поскольку они позволяют согласованно извлекать признаки из разных источников и повышать переносимость моделей на новые проекты[8 9]. Это важно для задач, где один тип данных дает неполную картину,

например при поиске мишеней и объяснении механизма действия, либо при выборе биомаркеров, где требуется связать молекулярный уровень с клиническими проявлениями.

Отдельным драйвером развития стала структурная революция в вычислительной биологии, которая напрямую влияет на продукты для discovery. Публикация AlphaFold 3 в Nature (2024)[10] закрепила тренд на более универсальное предсказание структурных комплексов и взаимодействий, включая случаи, когда в систему входят разные типы биомолекул[11]. Для практики разработки лекарств это означает рост качества и скорости этапов, связанных с построением гипотез о связывании, рациональном дизайне и приоритизации серий молекул. Важная особенность последнего периода развития состоит в том, что развитие идет не только в сторону повышения точности, но и в сторону применимости, так как модели все чаще встраиваются в продуктовые контуры, где результатом является не абстрактное предсказание, а ранжирование кандидатов, рекомендации по модификации структуры и более информированное планирование экспериментов.

Еще одним устойчивым глобальным трендом в рамках разработки лекарств является сближение ИИ-моделей с экспериментальной проверкой через замкнутые циклы обучения. Речь идет о подходах, где модель не просто оценивает кандидатов, но и помогает выбирать следующий набор экспериментов, а результаты лабораторных тестов используются для последующего обновления модели. В обзорах по самоуправляемым лабораториям (self-driving laboratories / SDL) подчеркивается, что закрытый контур «планирование эксперимента – выполнение – обновление модели» становится ключевым способом ускорения итераций в молекулярном поиске и оптимизации, особенно когда эксперимент дорогой, а пространство вариантов велико[12].
В контексте фармразработки это проявляется как рост числа продуктовых решений, ориентированных на ускорение цикла «проектирование – создание – тестирование» (design – make – test / DMT) и повышение доли подтвержденных экспериментов на ранних стадиях.

На этапе preclinical фокус ИИ-продуктов заметно смещается в сторону более комплексного прогнозирования свойств и рисков, он не остается на отдельных параметрах. В центре внимания остаются ADMET (абсорбция, распределение, метаболизм, выведение, токсичность лекарственного вещества) и безопасность, но глобально растет доля решений, которые комбинируют разные источники сигналов и дают системное представление о профиле кандидата, чтобы снизить вероятность позднего отказа[13]. Мультимодальные модели и предобучение на крупных наборах данных рассматриваются как один из наиболее перспективных путей для повышения устойчивости таких предсказаний и переноса между химическими классами и показаниями. Одновременно возрастает роль протоколов проверки, решения оцениваются не только по метрикам на отложенных наборах, но и по воспроизводимости при переносе на новые серии, лаборатории и режимы эксперимента, что становится де-факто ожиданием рынка к продуктам, претендующим на регулярное использование в доклиническом контуре.

На этапе clinical в 2024–2026 гг. усиливается тренд на применение ИИ для более точной работы с гетерогенностью пациентов и исходов. Наиболее

устойчивые направления связаны со стратификацией пациентов, выявлением биомаркеров и более информированным дизайном исследований, где ИИ используется для прогнозов ответа на терапию и формирования более однородных подгрупп, что потенциально повышает чувствительность исследования к эффекту[14]. Обзорные работы по AI в клинических исследованиях фиксируют рост применения ML/AI для поддержки решений в дизайне и анализе клинических испытаний, включая задачи подбора популяции, прогнозирования исходов и оптимизации аналитических планов при сложной структуре данных[15]. При этом ключевым критерием внедрения становится качество доказательной базы, клинические сценарии требуют высокой точности и ясной методологии валидации, а также устойчивости к дрейфу данных и различиям между центрами и когортами.

Наконец, сквозным трендом для всех трех стадий разработки является усиление требований к данным и доказуемости результата. Глобальная практика все чаще исходит из того, что качество входных данных, прослеживаемость происхождения, корректная разметка и репрезентативность выборок становятся не вспомогательными, а определяющими факторами эффективности ИИ-продуктов. В отраслевых и академических обзорах мультимодальный сдвиг прямо связывают с необходимостью управлять неоднородностью данных и обеспечивать воспроизводимость выводов, иначе рост сложности моделей не приводит к росту практической ценности. Поэтому наиболее заметные технологические траектории 2024–2026 гг. Объединяет тренд: ИИ в разработке лекарств развивается в сторону более универсальных и масштабируемых моделей, более тесной связи с экспериментальной проверкой и более строгих практик оценки результата, чтобы продукты можно было использовать как регулярный инструмент принятия решений на этапах поиска, доклиники и клиники. IMDRF N93 вводит понятие «алгоритмических рисков», которые возникают даже при формально корректной работе модели. К ним относятся проблема черного ящика (невозможность объяснить вывод), automation bias (чрезмерное доверие врача к ИИ), deskilling (деградация собственных клинических навыков), а также вредоносное вмешательство в данные обучения или в процесс обучения. Эти риски требуют специальных мер контроля, выходящих за рамки традиционного риск-менеджмента.

На этапе discovery основной класс ИИ-продуктов в мире сегодня связан с молекулярным представлением и генерацией, то есть с тем, как система интерпретирует химическую структуру и предлагает новые варианты, пригодные для дальнейшей оптимизации. На практике это реализуется как модели молекулярных представлений (включая 3D-подходы), которые затем используются в виртуальном скрининге, ранжировании серий и дизайне молекул de novo с заданными ограничениями по активности и свойствам[16 17 18 19 20]. Типовой результат таких продуктов для исследовательской команды — это упорядоченный список кандидатов, набор рекомендуемых модификаций (замены фрагментов, перестройки заместителей), а также оценка синтетической реализуемости и рисков по ключевым свойствам на раннем шаге.

Отдельный слой продуктов относится к задачам таргетинга, где ИИ применяется для выбора и обоснования терапевтических мишеней и для предсказани я взаимодействий «мишень – лиганд». Здесь важны решения,

которые интегрируют multi-omics-сигналы, сетевые модели биологии и исторические данные о таргетах, формируя приоритизацию и объяснимые гипотезы о механизме действия[21] [22]. В рамках того же класса находятся продукты для предсказания взаимодействия «лекарство – мишень» (drug – target interaction / DTI) и оценки вероятности связывания, которые используются как фильтр до затратных экспериментов и как инструмент ускорения перехода от биологической гипотезы к химическим сериям[23].

Для виртуального скрининга и оптимизации лидов (lead optimization) мировая линейка ИИ-продуктов строится вокруг графовых нейросетей и гибридов ML и докинга, где ML-модель помогает ранжировать позы связывания и повышать эффективность отбора кандидатов на эксперимент[24] [25]. Практически это выражается в появлении решений, которые дают воспроизводимое ранжирование и позволяют быстрее сужать химическое пространство при ограниченном бюджете синтеза и тестирования, особенно на стадиях hit-to-lead и lead optimization, когда приходится обрабатывать и анализировать множество (несколько десятков) аналогов, а не на миллионы соединений.

В сегменте биологических препаратов на этапе discovery один из наиболее заметных глобальных классов продуктов — генеративный дизайн белков, включая дизайн каркасов, биндеров и антителоподобных молекул. Современные решения используют диффузионные или близкие генеративные подходы и ориентированы на практический выход в виде кандидатов с заданной геометрией взаимодействия, предполагаемой стабильностью и технологичностью[26] [27]. С точки зрения процесса разработки лекарств важен результат, библиотека кандидатов, удовлетворяющих структурным ограничениям, и набор приоритетов для синтеза и биофизической валидации, что позволяет сократить число экспериментальных итераций в ранней оптимизации.

На этапе preclinical ключевые ИИ-продукты в мировом контуре разработки сосредоточены вокруг предсказания ADMET, PK/PD и рисков безопасности. Наиболее практичный формат таких решений — это платформы и сервисы, которые по структуре молекулы (и иногда по дополнительным экспериментальным признакам) возвращают профиль свойств и вероятностные оценки риска по нескольким параметрам сразу, обеспечивая раннее ограничение заведомо слабых кандидатов[28] [29]. Глобальная практика начинает чаще опираться на сопоставимые бенчмарки и наборы задач, которые позволяют сравнивать модели по единой рамке и избегать локально подходящих под частные случаи метрик без переносимости на новые серии.

Внутри доклиники заметен рост инженерных продуктов, которые автоматизируют построение и обновление предикторов, включая AutoML-подходы и стандартизированные пайплайны для оценки качества, что повышает воспроизводимость и снижает порог внедрения для команд, не являющихся ML-центрами. Для пользователя в R&D это выражается в том, что продукт становится не единичной моделью под одну задачу, а набором предикторов и процедур валидации, который можно регулярно применять при смене химических серий и терапевтических направлений, сохраняя сопоставимость решений между проектами[30].

Для этапа clinical глобальная линейка ИИ-продуктов, относящихся к разработке лекарства, концентрируется на трех задачах: стратификация

пациентов, работа с критериями включения и исключения и поддержка дизайна исследования, а также поиск и уточнение биомаркеров. В части стратификации и подбора пациентов быстро развивается класс решений, использующих большие языковые модели и NLP для сопоставления профиля пациента с описанием протокола и критериями отбора, что на выходе дает ранжирование потенциально подходящих исследований и объяснения по критериям[31 32]. Отдельно выделяются продукты для структурирования и оптимизации eligibility criteria (как части протокола), где данные электронных медицинских записей и NLP используются для оценки того, как изменения формулировок критериев влияют на доступную популяцию и реализуемость набора.

В задачах клинической разработки растет роль решений, которые связывают молекулярные данные, изображения и клинические параметры для поиска биомаркеров и определения подгрупп, где вероятность ответа на терапию выше. В мировой практике это проявляется как развитие инструментов исследования биомаркеров на мультиомных и клинических данных, а также как усиление роли вычислительной патологии в исследованиях (цифровые препараты, количественные признаки на частично ограниченных изображениях), где ИИ помогает извлекать стандартизируемые измеримые признаки, используемые в аналитике исследований[33 34 35]. Продуктовый выход для клинической команды здесь обычно формулируется как биомаркер-кандидат или набор признаков, определяющий подгруппы пациентов, а также модель, позволяющая воспроизводимо применять этот признак в последующих исследованиях или когортах.

Ключевых игроков глобального ландшафта ИИ-продуктов для разработки лекарств можно описать через несколько устойчивых групп, различающихся ролью в цепочке исследований, доклинической и клинической разработки. Первая группа — это крупные фармацевтические компании, которые одновременно развивают внутренние ИИ-платформы и системно выстраивают внешние технологические коллаборации под приоритетные R&D-направления. На практике это проявляется как публикации о внедрении ИИ в исследования и разработку, а также как партнерства, направленные на ускорение поиска мишеней и дизайн кандидатов на ранних стадиях[36 37]. Для отдельных компаний характерен акцент на собственные технологические решения для молекулярного дизайна, сопровождаемый публичной демонстрацией подходов и внутренних разработок, ориентированных на задачи drug discovery[38]. В биотехнологическом сегменте, тесно связанном с крупной фармой, заметны примеры развития следующего поколения внутренних AI-платформ для ускорения исследования и разработки лекарств, в том числе в сотрудничестве с технологическими партнерами по вычислениям и моделям[39 40]. На клинической границе разработки важную роль начинают играть и решения в области компаньонной диагностики, где ИИ применяется для поддержки выбора пациентов и биомаркеров, что отражается в публично заявляемых продуктах и регуляторных статусах для диагностических инструментов[41].

Вторая группа включает AI-native-компании, для которых ИИ является ядром продуктовой линейки в разработке лекарств. Их профиль обычно сочетает несколько типов продуктов: генеративный дизайн молекул, приоритизацию кандидатов, поддержку циклов «проектирование – создание – тестирование», а

также формирование собственных или совместных портфелей программ. Такие игроки подтверждают зрелость подхода через наличие собственных пайплайнов и постоянных технологических обновлений платформ, а также через многосторонние коллаборации с фармкомпаниями и биотехами, встроенные в реальную R&D-практику[42 43]. Внутри этой же группы выделяются компании, развивающие индустриальные платформы для совместной работы исследовательских команд, где продуктом становится единая среда для объединения экспериментальных данных, вычислительных предсказаний и командной коллаборации в ходе оптимизации кандидатов[44]. Отдельный подтип AI-native-игроков — это компании, комбинирующие ИИ, физическое моделирование и автоматизацию эксперимента и позиционирующие свои решения как ускорители исследования и оптимизации на ранних этапах разработки[45 46].

Третья группа — это технологические лидеры, формирующие базовые подходы и планку развития для отрасли, а также специализированные структуры, которые делают ставку на перенос достижений фундаментального ИИ в прикладной контур drug discovery. В этой логике важны не отдельные инструменты, а способность выстраивать масштабируемую R&D-модель, где результаты в одном проекте повышают эффективность последующих, а платформа поддерживает работу со сложными мишенями и классами терапевтиков. Партнерства такого типа, а также их расширения по мере накопления результатов отражают тот факт, что сфера рассматривает эти платформы как долгосрочный компонент discovery-стека[47 48 49]. Дополнительным маркером актуальности направления является расширение пула крупных фармпартнеров, что публично фиксируется отраслевыми медиа[50].

Четвертая группа — это провайдеры данных и инфраструктуры, которые обеспечивают производственный контур для ИИ в клинической разработке и на стыке клиники с биомаркерами. В глобальной практике клинические решения завязаны на доступ к высококачественным клиническим данным, данным реальной клинической практики и мультиомным наборам, а также на инструменты, позволяющие применять ИИ для задач подбора пациентов, поиска биомаркеров и поддержки клинических исследований. В этой группе заметны крупные провайдеры клинических сервисов и аналитики, которые развивают специализированные AI-инструменты и агенты для жизненного цикла клинических исследований и принятия решений в разработке[51 52]. Параллельно растет роль компаний, формирующих онкологические и мультиомные датасеты и решения для биомаркеров и клинического сопоставления, а также платформ real-world-данных, которые расширяют исследовательские возможности на стыке клиники и разработки препаратов[53 54]. Наконец, облачная инфраструктура и отраслевые платформы вычислений становятся стандартной основой для масштабирования моделей, обучения и безопасной совместной работы с данными, что отражается в профильных отраслевых программах и треках, ориентированных на клинические исследования[55].

1.2. Сегменты и динамика роста мирового рынка

Мировой рынок ИИ-продуктов для разработки лекарственных препаратов целесообразно рассматривать в разрезе двух крупных сегментов, динамика которых формируется разными типами задач и данных. Первый сегмент, AI-driven drug discovery и drug design, объединяет решения для генерации, отбора и оптимизации кандидатов, а также прогнозирования свойств и рисков (включая ADMET и безопасность) на ранних стадиях, где ключевым результатом являются более быстрые и качественные циклы выбора и доработки молекул. Второй сегмент включает решения для клинического контура, ориентированные на поддержку проектирования исследований, набора пациентов и мониторинга, где рост спроса связан с масштабированием клинических данных, усложнением протоколов и необходимостью повышать эффективность клинической разработки за счет аналитики и автоматизации процессов принятия решений.

Сегмент AI-driven drug discovery и drug design объединяет ИИ-продукты, которые применяются на ранних стадиях разработки лекарств для генерации и отбора кандидатов, их оптимизации и прогнозирования ключевых свойств. В отличие от решений клинического контура, ценность технологии чаще всего проявляется в качестве научно-исследовательских решений и скорости итераций, насколько быстрее и точнее команда приходит к набору перспективных молекул с приемлемым профилем свойств и рисков.

Сегмент, по оценкам разных аналитических источников, уже имеет самостоятельный заметный объем и высокую динамику, но разброс оценок остается существенным из-за различий в границах сегмента. В качестве ориентиров возможно привести несколько оценок. Например, Arizton Advisory & Intelligence оценивают объем так: порядка 1,71 млрд долл. в 2024 году с ростом до 8,52 млрд долл. к 2030 году[56]. Другие источники оценивают рынок выше. Например, Global Market Insights Inc. приводит данные: около 3,6 млрд долл. в 2024-м с прогнозом дальнейшего ускоренного роста в последующие годы[57]. При более консервативном определении сегмента также встречаются оценки около 1,5 млрд долл. в 2023 году с ростом до порядка 9,1 млрд долл. к 2030-му[58]. В совокупности эти оценки подтверждают, что сегмент уже сформирован и растет двузначными темпами, однако сопоставимость цифр зависит от закрепления границ и состава решений.

По зрелости сегмент относится к поздней стадии становления, технология перешла от единичных демонстраций к устойчивой практике, но рынок еще не полностью стандартизирован по метрикам результата и протоколам валидации. Подчеркивается два одновременно наблюдаемых факта: с одной стороны, быстрый прогресс генеративных моделей и моделей представлений для задач генерации, отбора и оптимизации кандидатов; с другой стороны, сохраняющиеся ограничения в переносимости результатов между проектами и зависимость качества от данных, дизайна эксперимента и корректности валидации[59 60]. Дополнительным индикатором зрелости является то, что значимая часть исследований и инструментов смещается к воспроизводимым продуктовым контурам — например, платформам ускоренного виртуального скрининга и ранжирования, где успех измеряется улучшением качества решений и скорости итераций в реальном R&D-процессе.

Спрос носит преимущественно корпоративный и проектно-портфельный характер, основными заказчиками выступают крупные фармкомпании и зрелые биотехи, которые покупают воспроизводимый контур ускорения принятия решений в раннем R&D (приоритизация мишеней, генерация и оптимизация кандидатов, снижение числа неудачных итераций), поэтому наиболее востребованы платформы и партнерские форматы, встроенные в существующие процессы discovery и preclinical. При этом спрос усиливается по мере наращивания собственных данных и вычислительных мощностей у фармы, что отражается в крупных технологических партнерствах, например: совместные вычислительные инфраструктуры для обучения моделей на экспериментальных данных, инициативы совместного обучения моделей на объединенных промышленных датасетах при сохранении конфиденциальности. Это указывает на переход рынка к индустриальному уровню применения ИИ и ориентацию спроса на масштабируемые решения.

Сегмент решений для клинического контура включает ИИ-продукты, которые применяются в клинической разработке для поддержки проектирования исследований, подбора и удержания пациентов, выбора центров, мониторинга качества и безопасности, а также для более эффективной работы с растущими массивами клинических и цифровых данных. По оценкам аналитических источников, сегмент уже имеет самостоятельный масштаб. Например, MarketsandMarkets оценивает рынок AI in clinical trials в 1,35 млрд долл. в 2024 году с прогнозом 2,75 млрд долл. к 2030-му[61]. Другие оценки дают более высокие траектории роста. Например, Mordor Intelligence оценивает рынок в 2,68 млрд долл. в 2026-м с прогнозом 8,24 млрд долл. к 2031 году, что отражает более широкие границы сегмента и включение большего числа функциональных сценариев[62]. В близком контуре отдельные провайдерские оценки (AI-based clinical trials solution provider) также указывают на рынок порядка 2,37 млрд долл. в 2024 году, подтверждая наличие сформированного спроса и самостоятельной продуктовой экосистемы[63].

По зрелости этот сегмент заметно продвинулся к промышленной фазе за счет стандартизации процессов клинических исследований и усиления регуляторной определенности вокруг цифровых и удаленных элементов, но при этом остается неоднородным по уровню доказуемости эффекта в разных функциях. Обновление международного стандарта надлежащей клинической практики ICH E6(R3) закрепляет фокус на рискориентированном управлении качеством, планах мониторинга и управлении данными, что повышает спрос на инструменты предиктивной аналитики, автоматизированного контроля качества и структурирования данных[64]. Дополнительно зрелость сегмента поддерживается тем, что регуляторные рекомендации по децентрализованным элементам клинических исследований формализуют практики удаленного взаимодействия и цифрового сбора данных, увеличивая объем данных и стимулируя применение аналитических и ИИ-инструментов для мониторинга и управления исследованиями. Одновременно академические обзоры фиксируют расширение спектра применений ИИ по жизненному циклу клинического исследования от дизайна и рекрутинга до мониторинга и анализа, но подчеркивают, что ключевым ограничением остаются переносимость решений, качество данных и корректность валидации, поэтому зрелость сегмента следует

понимать как высокую по широте внедрений, но неравномерную по уровню доказательной стандартизации в разных типах продуктов.

Спрос на ИИ-решения для клинического контура формируется прежде всего со стороны спонсоров клинических разработок и операционных исполнителей, таких как фармацевтические компании, биотехи, и контрактных исследовательских организаций, для которых ключевыми болевыми точками остаются сроки и реализуемость набора пациентов, качество данных, стоимость мониторинга и риск отклонений от протокола. Поэтому наиболее востребованы продукты, которые дают измеримый операционный эффект в конкретных функциях клинического исследования, такие как ускорение идентификации и привлечения подходящих пациентов, оптимизация подбора и поддержки исследовательских центров, автоматизация и рискориентированный мониторинг, а также инструменты для структурирования и анализа данных из разнородных источников, включая EHR и цифровые сигналы[65]. В 2024–2026 гг. спрос дополнительно усиливается за счет распространения децентрализованных элементов исследований и роста объема цифровых данных, что повышает потребность в аналитических и ML-инструментах для контроля качества и безопасности в реальном времени, а также в масштабируемых платформах управления клиническими данными и процессами.

1.3. Драйверы и ограничения развития рынка

Рост мирового рынка ускоряется одновременно за счет давления сверху со стороны экономики R&D и выталкивания снизу со стороны технологической готовности. С одной стороны, фарма и биотех сталкиваются с устойчивым ростом сложности биологии, портфелей и высокой ценой ошибок на ранних этапах, поэтому растет спрос на инструменты, которые улучшают качество отбора кандидатов и снижают число неэффективных итераций в discovery и доклинической оптимизации, а в клинике повышают реализуемость исследований через более точную работу с популяциями, данными и рисками. С другой стороны, технологический прогресс последних лет сделал ИИ более применимым как продукт, генеративные и фундаментальные модели расширили спектр задач от предсказания свойства к проектированию кандидатов и поддержке принятия решений, мультимодальные подходы позволяют использовать одновременно химические, биологические и клинические данные, а достижения в структурной биологии и вычислительной патологии усиливают доказательную основу биомаркерных и стратификационных решений. Одновременно растет доступность инфраструктуры и данных. Инвестиции в высокопроизводительные вычисления, переход к промышленным облачным и on-prem-решениям, а также запуск масштабных датасетов и совместных инициатив повышают скорость обучения моделей и делают внедрение возможным не только для узкой группы лидеров, но и для более широкого круга компаний. Клинический контур дополнительно поддерживается стандартизацией процессов качества и расширением децентрализованных элементов исследований, что увеличивает объем цифровых данных и создает естественный спрос на инструменты автоматизированного мониторинга, контроля качества и аналитики в реальном времени.

Масштабирование рынка ограничивается прежде всего тем, что эффект ИИ трудно воспроизводимо перенести между проектами, площадками и популяциями при неизбежной неоднородности биологических и клинических данных. В discovery и доклинической разработке ключевой барьер связан с качеством и репрезентативностью экспериментальных данных, разные лабораторные протоколы, различия в условиях эксперимента, смещение распределений при переходе к новым химическим сериям и терапевтическим мишеням приводят к тому, что модели часто демонстрируют приемлемое качество внутри распределения, но существенно теряют точность на новых задачах, где требуется реальная ценность. Эффект усиливается тем, что значительная доля релевантных данных является закрытой, фрагментированной и ограниченной юридически как в части IP, так и в части доступа и передачи, поэтому обучение и валидация часто происходят на неполных и несопоставимых выборках. В результате компании вынуждены тратить значимые ресурсы на построение промышленного контура данных, согласование форматов, процедур разметки и качества, иначе модели становятся локальными инструментами, пригодными только в пределах одного проекта или команды.

В клиническом контуре ограничения носят операционно-регуляторный характер и связаны с тем, что клинические данные структурно более неоднородны и подвержены системным проблемам качества: пропуски и несогласованность в EHR, различия между центрами, смена практик кодирования, а также быстрое появление новых цифровых источников данных при децентрализованных элементах исследований. Даже высокоточные модели могут деградировать со временем из-за дрейфа данных и изменения контекста применения, что требует постоянного мониторинга, повторной валидации и процедур жизненного цикла модели, иначе возрастает риск некорректных решений в ходе исследования. Усиливается нагрузка на доказуемость и прозрачность, для продуктов, влияющих на решения в клинической разработке, требуется воспроизводимая документация, контроль версий, обоснование применимости в конкретном контексте и соблюдение требований надлежащей клинической практики и управления качеством. Дополнительный слой ограничений создают требования к безопасности и комплаенсу при применении AI/ML в медицинском и клиническом контексте, включая нормы к прозрачности, управлению рисками, технической документации и ответственности, которые повышают стоимость внедрения и удлиняют цикл от пилота масштабирования.

Сквозным барьером масштабирования по всему рынку остается разрыв между технологической демонстрацией и промышленной эксплуатацией. Даже когда модель показывает результат, масштабирование требует интеграции в рабочие процессы, распределения ответственности за решения, подготовки персонала и устойчивого контура эксплуатации (обновление данных, контроль качества, управление дрейфом и инцидентами). На уровне отрасли это приводит к тому, что спрос смещается от моделей к платформенным решениям и консорциумам, которые позволяют безопасно работать с закрытыми данными и выстраивать совместимые практики валидации. Однако такие форматы сложны, они требуют согласования прав на данные, технических механизмов приватности, единых протоколов качества и доверия между участниками, а также долгосрочных организационных инвестиций. В совокупности эти факторы

объясняют, почему рынок растет быстро, но масштабирование остается неравномерным, в задачах, где данные стандартизируемы и эффект можно измерить процессными метриками, внедрение ускоряется, тогда как в задачах с высокой неоднородностью данных и высокой ценой ошибки развитие идет через более строгие контуры валидации и более медленное промышленное распространение. Один из ключевых разрывов — кадровый: выпускники не владеют системным мышлением в области разработки лекарств, не знакомы с GxP-требованиями, проектным управлением и анализом рисков. Для преодоления разрыва целесообразно корректировать ФГОС специалитета по специальности «Фармация», открывать программы «провизор-исследователь», внедрять бизнес-инструменты и системы менеджмента качества в университетах, а также запускать совместные с фармкомпаниями магистратуры и ДПО[66 67].

Факторы, снимающие ограничения масштабирования во многом связаны с индустриализацией и стандартизацией данных и процессов, то есть с переводом ИИ из режима отдельной модели в режим управляемого производственного контура. Ключевое условие здесь заключается в стандартизации данных и воспроизводимых правилах их подготовки, в клиническом контуре это поддерживается широким внедрением общепринятых стандартов структурирования и трассируемости клинических данных, что делает возможным повторное использование данных, сопоставимость результатов и более быстрые циклы анализа. В дополнение к стандартам критическим фактором становится движение к единым моделям представления и терминологиям для наблюдательных данных, позволяющим объединять разнородные базы и воспроизводимо переносить аналитические процедуры между организациями. Параллельно усиливается практическая ориентация на принципы качества данных и метаданных, когда данные проектируются для машинного использования с точки зрения выявления, интероперабельности и повторного применения. В результате снижается доля «ручной» работы по подготовке данных под каждый отдельный проект и растет переносимость ИИ-решений между программами разработки.

Второй набор факторов связан с формализацией практик разработки и эксплуатации моделей, которые позволяют безопасно использовать ИИ в задачах с высокой ценой ошибки. В клинической разработке этому способствует закрепление рискориентированных подходов к качеству и управлению данными и процессами, включая более четкие ожидания к документации, контролю версий, мониторингу качества и управлению изменениями, что упрощает встроенное применение аналитических и ML-инструментов. Одновременно развиваются рамки для жизненного цикла AI/ML-систем, ориентированные на надежность, управляемость рисков и качество, и они начинают играть роль общего языка между разработчиками технологий, клиническими командами и функциями качества. Практический эффект состоит в том, что масштабирование перестает быть серией уникальных внедрений вручную и превращается в повторяемый процесс, состоящий из определения контекста применения, закрепления метрик качества и безопасности, описания процедуры обновления и мониторинга, а затем тиражирования контура на новые исследования или терапевтические области.

Третий фактор, снимающий ограничения, связан с механизмами совместной работы с закрытыми и чувствительными данными без прямой передачи исходных массивов между организациями. Для отрасли разработки лекарств это особенно важно, потому что значимая часть наиболее ценных данных защищена коммерчески и юридически, а клинические данные дополнительно ограничены требованиями к приватности. Рост практик федеративного анализа и других сохраняющих приватность подходов позволяет расширять обучающие выборки и повышать устойчивость моделей, не разрушая рамки конфиденциальности и комплаенса. Это создает условия, при которых отрасль может одновременно увеличивать объем доступного учебного материала для моделей и сохранять доверие к процессам работы с данными, что в долгосрочной перспективе поддерживает более равномерное масштабирование рынка. В совокупности устойчивый рост будет обеспечиваться там, где одновременно выполняются три условия: данные стандартизированы и пригодны для повторного использования, жизненный цикл моделей управляем и проверяем, а обмен знаниями и улучшение моделей возможны в форматах, совместимых с требованиями конфиденциальности и качества.

1.4. Правовое регулирование на различных рынках

Правовое регулирование применения ИИ в фармацевтической отрасли находится на стадии активного формирования и существенно различается по юрисдикциям с точки зрения степени детализации, обязательности требований и охвата жизненного цикла лекарственного препарата. На крупнейших рынках складывается многоуровневый контур регулирования, включающий как общие акты в сфере ИИ и данных, так и специализированные документы для разработки, оценки и регистрации лекарственных препаратов, что постепенно повышает требования к прозрачности моделей, качеству данных, воспроизводимости результатов и доказуемости их использования в регуляторных и R&D-процессах.

Гармонизация регуляторных подходов к применению моделей и алгоритмов в разработке лекарственных препаратов усиливается за счет принятия ICH M15, который формирует общую международную рамку для модельноориентированной разработки лекарств. Документ закрепляет следующее: доказательства, полученные с использованием вычислительного моделирования и симуляции, могут применяться для поддержки решений на различных этапах разработки препарата, а в перечень релевантных методов прямо включаются ИИ и МО. Тем самым AI/ML переводятся из категории экспериментальных инструментов в контур формализованного регуляторного рассмотрения при условии их надлежащего планирования, оценки и документирования[68]. К 2023 году гармонизировано 10 методических руководств ICH, а также внедрено более 30 региональных руководств по искусственному интеллекту (ИИ), машинному обучению (МО) и математическому моделированию (ММ): в США — 23, в ЕС — 7, в КНР — 2, в Японии — 3, в Канаде — 1; бразильский регулятор и регулятор Саудовской Аравии также разрабатывают подобные документы[69] [70].

Практическая значимость ICH M15 для рынка заключается в унификации ожиданий регуляторов в отношении контекста использования, влияния моделей и последствий неверного решения. Такой подход повышает требования к воспроизводимости результатов, качеству исходных данных, прозрачности логики модели, полноте model evaluation и качеству отчетности для регуляторных взаимодействий. Для разработчиков это означает рост затрат на методологическую и документарную подготовку AI/ML-решений, но одновременно снижает неопределенность в отношении того, каким образом модельные результаты могут быть интегрированы в регуляторный пакет и использованы при обосновании решений по разработке препарата[71].

С точки зрения рыночной динамики ICH M15 выступает одновременно драйвером и ограничением. С одной стороны, документ повышает институциональную легитимность AI/ML и других подходов в фармацевтическом R&D, создавая более предсказуемую среду для их внедрения в международных программах разработки. С другой стороны, закрепление единых ожиданий к планированию, отчетам об анализе модели, документации для подачи заявок и взаимодействию с регуляторами повышает порог входа для компаний, не обладающих зрелыми компетенциями в области управления данными, проверки моделей и регуляторного сопровождения цифровых инструментов.

В США формируется специализированный регуляторный контур для применения ИИ в разработке лекарственных и биологических препаратов. Проект, опубликованный в январе 2025 года, распространяется на случаи, когда модели используются для получения данных или информации, предназначенных для поддержки регуляторных решений в отношении безопасности, эффективности или качества препарата. Документ охватывает неклинический, клинический, пострегистрационный и производственный этапы жизненного цикла продукта, при этом не распространяется на ранний исследовательский контур и на внутренние операционные сценарии, не влияющие на безопасность пациента, качество продукта или надежность результатов исследований[72]. IMDRF N93 построен вокруг логики ISO 14971 (риск-менеджмент). Все этапы жизненного цикла ИИ-медизделия от планирования до вывода из эксплуатации рассматриваются через призму идентификации, анализа, контроля и мониторинга рисков. Документ подчеркивает, что базовые принципы управления рисками для медицинских изделий остаются неизменными, но требуют адаптации к специфическим ИИ-рискам. Помимо руководства по ИИ, FDA использует разветвленную систему ускоренных процедур, такие как Priority Review (сокращение срока рассмотрения до 6 месяцев), Accelerated Approval (по суррогатным конечным точкам), Breakthrough Therapy (интенсивное сопровождение), Orphan Drug Designation (налоговые кредиты, 7 лет эксклюзивности). Для малых молекул действует NDA, для биотехнологических — BLA[73].

Ключевое значение руководства заключается во внедрении рискориентированной модели оценки достоверности AI-систем через оценку достоверности структуры, привязанной к конкретному случаю использования. Подход FDA предполагает, что уровень требований к валидации, документированию, качеству данных, мониторингу и объяснимости модели должен соотноситься с тем, насколько сильно результаты AI влияют на

регуляторное решение и каковы последствия потенциальной ошибки. Для рынка это означает переход от экспериментального использования AI к более формализованному режиму регуляторного рассмотрения. С одной стороны, снижается неопределенность в отношении приемлемости AI-инструментов, с другой — повышаются требования к управлению данными, оценке моделей, отслеживаемости и качеству доказательной базы на всем жизненном цикле применения модели[74].

В Европейском союзе регулирование применения ИИ в фармацевтической отрасли формируется как сочетание общих норм в сфере ИИ и специальных подходов, ориентированных на жизненный цикл лекарственного препарата. Одним из ключевых профильных документов является Reflection paper on the use of artificial intelligence in the medicinal product lifecycle, опубликованный Европейским агентством по лекарственным средствам. Документ рассматривает применение ИИ на этапах разработки, регистрации, производства и пострегистрационного обращения препаратов и фиксирует, что использование таких решений должно оцениваться с учетом конкретной задачи, характеристик данных, степени влияния системы на итоговое решение и возможности надлежащего контроля со стороны человека. EMA также применяет механизмы ускоренного доступа, такие как Accelerated Assessment (сокращение сроков оценки CHMP), PRIME (ранняя поддержка и право на ускоренную оценку), Conditional Marketing Authorisation (на основе неполных данных), Exceptional Circumstances, Orphan Medicinal Product designation. С 2022 года действует единая платформа CTIS для подачи и оценки заявок на клинические испытания[75].

Практическое значение данного подхода заключается в том, что европейский регулятор связывает допустимость применения ИИ не с самой технологией, а с качеством ее внедрения и обоснования. В центре внимания находятся надежность и пригодность модели, полнота документации, прослеживаемость данных и результатов, соблюдение требований законодательства о персональных данных и наличие механизмов управления рисками на всем протяжении использования системы. Для рынка это означает постепенный переход к более формализованному порядку оценки таких решений. С одной стороны, создаются более понятные ориентиры для компаний, внедряющих ИИ в фармацевтические процессы, с другой — возрастают требования к качеству данных, внутреннему контролю и доказательству корректности работы модели в каждом конкретном сценарии применения[76].

Таким образом, на крупнейших фармацевтических рынках формируется переход от общего признания потенциала ИИ к его поэтапному включению в формализованный регуляторный контур. Несмотря на различия в подходах США и ЕС, общее направление заключается в повышении требований к качеству данных, прозрачности применяемых моделей, воспроизводимости результатов и обоснованности их использования в процессах разработки и вывода лекарственных препаратов на рынок.

1.5. Инвестиции, кейсы и стартапы на глобальных рынках

В 2024–2026 гг. самые показательные инвестиционные сделки концентрируются вокруг масштабирования AI-driven drug discovery и drug design через крупные многолетние партнерства. Типовая направленность таких сделок — получение доступа к генеративному дизайну малых молекул или белковых терапевтиков, ускорение циклов «проектирование – создание – тестирование» и расширение пайплайна за счет более быстрой приоритизации и оптимизации кандидатов. Параллельно усиливается тенденция к укрупнению и платформизации игроков, что проявляется в консолидационных сделках, рынок постепенно отходит от множества узких решений к более интегрированным end-to-end-платформам, объединяющим вычислительный дизайн, большие биоданные и экспериментальные контуры.

Растет пул сделок, которые поддерживают клиническую разработку инфраструктурой и данными. Это совместные вычислительные лаборатории и суперкомпьютерные инициативы, а также кооперации по закрытым промышленным датасетам, критичным для качества моделей взаимодействий и переносимости результатов. Показательный вектор последних лет — переход к совместному обучению и обмену знаниями без раскрытия исходных данных, в частности на федеративных платформах и консорциумах, что позволяет расширять обучающие выборки за пределами одной компании и повышать качество моделей для малых молекул и структурной биологии. В сумме эти сделки показывают, что средства вкладываются в сферы, где можно увеличить масштаб и воспроизводимость эффекта либо через платформы, которые стабильно поставляют кандидатов и ускоряют ранний R&D, либо через вычисления и данные, которые делают такие платформы промышленно применимыми (таблица 1.1).

Таблица 1.1 — Сделки на глобальном рынке использования ИИ в драгдизайне

Дата	Участники	Тип сделки	Оценка	Направленность
09.02.2026	Takeda – Iambic	Партнерство	потенциально более 1,7 млрд долл. и роялти	AI-дизайн малых молекул (онкология, гастровоспаление)
20.11.2025	Merck KGaA – Valo Health	Партнерство	потенциально более 3 млрд долл. и роялти	AI-платформа и human data для discovery (нейродегенерация, болезнь Паркинсона)
01.10.2025	Bristol Myers Squibb – Takeda – Astex и др. (консорциум)	Дата-кооперация/ консорциум	не раскрыто	Объединение закрытых структурных данных для обучения модели взаимодействий белок – малая молекула (OpenFold3)

Таблица 1.1 (окончание)

Дата	Участники	Тип сделки	Оценка	Направленность
12.01.2026	NVIDIA – Eli Lilly	Стратегическое партнерство, открытие совместной лаборатории	1 млрд долл. за 5 лет	Вычисления и совместная AI-лаборатория для ускорения drug discovery и development
31.03.2025	Isomorphic Labs – Thrive Capital (и др.)	Раунд финансирования	600 млн долл.	Масштабирование AI-платформы drug design и вывода кандидатов в разработку
08.08.2024	Recursion – Exscientia	M&A	688 млн долл.	Консолидация AI-платформ (биология и химия) для discovery и оптимизации кандидатов
14.10.2025	Takeda – Nabla Bio	Расширение партнерства	потенциально более 1 млрд долл.	AI-дизайн белковых терапевтиков для раннего пайплайна
13.01.2026	AstraZeneca – Modella AI	Приобретение	не раскрыто	AI-агенты для онкологических программ, поддержка клинической разработки и биомаркеров
09.01.2026	Schrödinger – Eli Lilly	Коммерческая коллаборация	не раскрыто	Дистрибуция и масштабирование AI-платформы Lilly через LiveDesign для ускорения discovery

В 2024–2026 гг. практики применения ИИ крупными фармацевтическими компаниями все чаще оформляются в долгосрочные программы, в которых ИИ становится частью базовой R&D-инфраструктуры. Это видно по переходу к совместным исследовательским центрам компетенций, где объединяются доменные эксперты фармкомпании, вычислительные команды и доступ к масштабным вычислениям, чтобы ускорять циклы разработки и повышать качество решений на ранних стадиях. Одновременно растет фокус на данных промышленного качества, большие компании участвуют в проектах по созданию крупных биологических датасетов, которые затем используются для обучения и улучшения моделей, ориентированных на поиск мишеней и раннее понимание биологии заболеваний. На уровне организации R&D это означает смещение от закупки инструмента к формированию постоянно работающего контура, где данные, модели и эксперименты взаимно усиливают друг друга, а эффект измеряется устойчивым улучшением качества выбора программ и скорости итераций.

Параллельно усиливается тенденция к межкорпоративной кооперации между самими крупными компаниями, когда конкурентные игроки объединяют закрытые данные или инфраструктуру ради повышения качества базовых моделей, которые затем применимы к широкому классу задач разработки. Показательный пример такого подхода — это консорциум, в котором крупные фармкомпании загружают структурные данные «белок – малая молекула» для

обучения модели взаимодействий, используя федеративную платформу, чтобы не раскрывать исходные наборы данных, что показывает переход отрасли к индустриальным механизмам совместного обучения и масштабирования качества. В клиническом контуре у крупных компаний заметна ориентация на ИИ как на средство повышения вероятности успеха клинических программ через более точный отбор пациентов и биомаркерную стратегию, что проявляется в приобретениях и расширении внутренних возможностей в количественной патологии и клинической аналитике. В сумме крупные компании демонстрируют, что использование ИИ смещается к двум устойчивым направлениям: построение долгоживущих платформ и центров и расширение доступа к данным и моделям через партнерства и консорциумы, чтобы повысить воспроизводимость эффекта в масштабах портфеля (таблица 1.2).

Таблица 1.2 — Кооперационные цепочки использования ИИ в фармацевтической инфраструктуре

Компания	Формат использования	Сегмент	Функции ИИ	Источник
Sanofi	Технологическое сотрудничество	Клиническая разработка	AI-софт для ускорения drug development, включая аналитику и поддержку клинических процессов	*https://www.sanofi.com/en/media-room/press-releases/2024/2024-05-21-05-30-00-2885244*
Pfizer	Расширение сотрудничества	Discovery / medicinal chemistry	Использование ИИ для поддержки медицинской химии	*https://cen.acs.org/business/start-ups/Pfizer-doubles-down-AI-partnership/103/web/2025/01*
Pfizer	Расширение стратегического сотрудничества	Discovery / preclinical	Усиление платформы молекулярного моделирования и предикторов для работы с маломолекулярным химическим пространством и ускорения отбора и оптимизации	*https://www.prnewswire.com/news-releases/xtalpi-and-pfizer-expand-strategic-collaboration-to-advance-ai-driven-drug-discovery-and-materials-science-simulations-302494033.html*
GSK	Индустриально-академическая программа	Discovery / ранняя биология	Применение advanced AI для ускорения исследований против антимикробной резистентности и приоритизации направлений/мишеней	*https://www.gsk.com/en-gb/media/press-releases/gsk-and-fleming-initiative-scientists-unite-to-target-amr-with-advanced-ai/*
Novartis	Внутреннее внедрение	Клиническая разработка	Автоматизация подготовки клинических отчетов и медицинской документации с использованием AI-технологий	*https://www.novartis.com/sites/novartiscom/files/novartis-q1-2024-impact-and-sustainability-update.pdf*

Среди стартапов в 2024–2026 гг. особенно выделяются компании, которые строят платформы следующего поколения вокруг генеративных моделей и собственной генерации данных. Xaira Therapeutics запустилась с заявленным объемом финансирования в 1 млрд долл. и акцентом на фундаментальные модели и масштабирование полного контура ранней разработки. Generate Biomedicines привлекает внимание как один из лидеров в генеративном дизайне белковых терапевтиков, что подтверждается крупными партнерствами, включая многотаргетное сотрудничество с Novartis, где есть существенные авансовые и потенциальные milestone-платежи. SandboxAQ демонстрирует отдельную линию развития сегмента, делая ставку на синтетические массивы данных для ускорения предсказания связывания «малая молекула – белок», что отражает тренд на масштабирование качества моделей за счет данных, а не только архитектур. Параллельно появляются подходы AI science factory, где стартапы объединяют специализированные модели и автоматизированные лабораторные контуры для непрерывной генерации экспериментов и данных.

В клиническом контуре наиболее интересны стартапы, которые либо уменьшают размер контрольных групп и повышают мощность исследований за счет цифровых двойников, либо радикально ускоряют рекрутинг и работу с разнородными клиническими данными. Unlearn развивает подход digital twins для формирования прогнозов исходов у участников контрольной группы и привлекла 50 млн долл. Series C, что иллюстрирует устойчивый спрос на методологические AI-инструменты клинической разработки. В сегменте рекрутинга и поиска пациентов заметен Deep 6 AI, чья технология для работы с EHR и ускорения подбора пациентов стала настолько значимой, что была приобретена Tempus, это можно рассматривать как признак зрелости направления AI-поддержки набора и клинических операций. Для задач клинического поиска и синтеза медицинских знаний в интересах разработки препаратов интересен Mendel, который продвигает архитектуры, ориентированные на снижение галлюцинаций в медицинских сценариях и масштабируемое применение в фарме через облачные каналы распространения (таблица 1.3).

Таблица 1.3 — Стартапы и инвестиционные проекты

Стартап	Страна	Фокус	Контур	Финансирование	Источник
Xaira Therapeutics	США	AI-first-платформа для разработки end-to-end, ставка на foundation models и масштабирование полного контура	Discovery, preclinical	запуск с 1 млрд долл. committed funding	*https://www.fiercebiotech.com/biotech/new-ai-drug-discovery-powerhouse-xaira-rises-1b-funding*
Generate: Biomedicines	США	Генеративный дизайн белковых терапевтиков	Discovery	> 1 млрд долл.	*https://generatebiomedicines.com/media-center/generatebiomedicines-announces-multi-target-collaboration-with-novartis; https://tto.dartmo*

Таблица 1.3 (окончание)

Стартап	Страна	Фокус	Контур	Финансирование	*Источник*
					uth.edu/news/2024/09/generatebiomedicines-announces-multi-target-collaboration-novartis-discover-and
SandboxAQ	США	Синтетические 3D-датасеты и модели для предсказания связывания «малая молекула — белок»	Discovery, preclinical	почти 1 млрд долл. венчурного капитала	*https://www.reuters.com/business/healthcare-pharmaceuticals/nvidia-backed-ai-startup-sandboxaq-creates-new-data-speed-up-drug-discovery-2025-06-18/*
Lila Sciences	США	Сочетание моделей и экспериментального контура для ускорения научных циклов	Discovery, preclinical	> 1,3 млрд долл.	*https://www.reuters.com/business/ai-lab-lila-sciences-tops-13-billion-valuation-with-new-nvidia-backing-2025-10-14/*
Unlearn	Великобритания	Digital twins для клинических исследований	Clinical	Series C 50 млн долл.	*https://www.businesswire.com/news/home/20240206806844/en/Unlearn-Raises-%2450-Million-Series-C-to-Optimize-Clinical-Research-With-AI-Powered-Digital-Twin-Technology*
Deep 6 AI	США	Поиск пациентов, когорт по EHR для ускорения рекрутинга	Clinical	–	*https://investors.tempus.com/news-releases/news-release-details/tempus-announces-acquisition-deep-6-ai*
Mendel	США	Клинический поиск и синтез знаний для фармы; акцент на снижение галлюцинаций в медсценариях	Clinical	–	*https://www.pharmavoice.com/news/mendel-ai-hallucination-hypercube-google-cloud/725290/*

ЧАСТЬ II
РОССИЙСКИЙ РЫНОК ИИ В ФАРМЕ И DRUG DISCOVERY

2. Российский рынок ИИ в фарме и Drug Discovery

2.1. Структура рынка и ключевые участники в России

Российский рынок ИИ в фарме формируется вокруг крупных производителей и держателей R&D-портфелей, которые выступают заказчиками решений и владельцами ключевых экспериментальных контуров. Также ключевыми участниками выступают институты и технологические партнеры, которые являются центрами компетенций и агрегируют экспертизу по моделям, данным и методологиям AI-drug discovery. Примером такого игрока в России является Институт AIRI, который публично участвует в отраслевых партнерствах и инициативах по разработке лекарств с применением ИИ[77].

Также на рынке функционируют AI-native компании и резиденты инновационных кластеров, которые либо создают собственные платформы, либо участвуют в совместных исследовательских программах. В биомедицинском кластере Сколково в качестве резидента есть российское подразделение Insilico Medicine, реализующее международные практики и подходы к AI-драгдизайну через локальную инфраструктуру[78].

По структуре спроса рынок пока преимущественно проектный: преобладают конкретные программы (под отдельную терапевтическую область или класс молекул), а не масштабные коробочные решения. Это связано с тем, что ценность создается не только алгоритмами, но и доступом к данным, качественной экспериментальной валидацией и возможностью быстро проводить циклы от гипотезы до обновления модели[79].

2.2. Уровень зрелости и текущие практики применения ИИ

По зрелости практик Россия в основном находится на уровне прикладных пилотов и ограниченных продуктовых контуров, где ИИ используется для ускорения ранних стадий R&D (выбор, оптимизация кандидатов, приоритизация направлений, поддержка анализа данных). Например, публично представленный проект по нейродегенеративным заболеваниям, где «Р-Фарм» и AIRI при поддержке Сбера заявляют использование генеративного ИИ для разработки инновационных препаратов[80].

Отдельно стоит выделить попытку перейти от точечных моделей к интеграции ИИ по этапам фармразработки. В частности, ПРОМОМЕД, Сбер и AIRI публично объявляли о сотрудничестве, при этом цель формулируется как совместная разработка и внедрение AI-решений на всех этапах создания лекарств[81].

Более зрелые и часто более измеримые применения возникают в производстве и инженерных процессах (оптимизация режимов, контроль параметров, анализ потоков данных от оборудования). В российских материалах и выступлениях встречаются примеры, где BIOCAD обсуждает применение ML в инженерных задачах, в том числе вокруг биореакторов и данных процесса, что типично для направления AI for manufacturing & quality[82]. На российском есть рынке есть ряд компаний, которые уже обладают рядом

цифровых компетенций в использовании ИИ в фармацевтике. Так, например, компания «СимургФарм» ведет разработку ИТ-продуктов для анализа биомедицинских данных, решения задач количественной фармакологии и оптимизации разработки лекарств. Основной продукт компании — вычислительная платформа «Симург»[83].

2.3. Инвестиции, партнерства и кейсы на российском рынке

В 2022–2026 гг. российские крупные сделки в применении ИИ для разработки лекарств чаще всего оформлялись как стратегические партнерства и совместные программы между крупной фармой, технологическими экосистемами и исследовательскими институтами, где ключевая ценность заключалась в доступе к вычислительной инфраструктуре, данным и компетенциям[84]. Наиболее заметный по масштабу участников трек связан с экосистемой Сбера и фармгруппой «Р-Фарм», в публичных материалах описывается запуск совместных AI-решений для ускорения разработки структуры молекулы, кандидата и развитие сотрудничества в направлении биологических препаратов и малых молекул, что отражает ориентацию российской практики на ускорение ранних стадий discovery и design через совместные модели и вычисления.

Параллельно формируется второй тип крупных инициатив, где сделка выражается в создании платформ и межорганизационных проектов вокруг ИИ-контуров разработки, например, проекты Сколтеха и НТУ «Сириус» по AI-платформе для ускорения поиска инновационных лекарств, а также расширение коллабораций «Р-Фарм» с исследовательскими институтами (AIRI) при поддержке крупного технологического партнера, в том числе в направлении таргетных антител и сложных систем доставки, что показывает смещение российского рынка к кооперационным моделям для задач R&D (таблица 2.1).

Таблица 2.1 — Сделки на российском рынке

Дата	Участники	Формат	Контур	Направленность	Источник
08.09.2022	Сколтех, МГУ, НТУ «Сириус»	Платформа	Discovery, preclinical	Платформа Syntelly для поиска и анализа новых соединений: прогноз токсикологических и физико-химических свойств, визуализация химпространства, прогноз синтеза	*https://pharmmedprom.ru/news/iskusstvennii-intellekt-dlya-sozdaniya-novih-lekarstv-razrabotali-skolteh-i-ntu-sirius/*
13.12.2024	Сбер, AIRI	Создание центра	Discovery, preclinical	Создание центра AIDD для разработки лекарств с использованием ИИ (поиск мишеней, создание новых молекул, доклинические исследования)	*https://pharmmedprom.ru/news/v-rossii-sozdadut-tsentr-razrabotki-lekarstv-s-pomoschyu-iskusstvennogo-intellekta/*

Таблица 2.1 (окончание)

Дата	Участники	Формат	Контур	Направленность	Источник
25.06.2025	Сбер, «Р-Фарм», AIRI	Стратегическое партнерство	Discovery, preclinical, clinical	Разработка ИИ-решений для ключевых процессов фармы, включая поиск перспективных молекул и разработку новых препаратов	*https://gxp-academy.org/news/partner-evraziyskoy-akademii-nadlezhashchikh-praktik-kompaniya-r-farm-zaklyuchila-soglashenie-o-stra/*
23.06.2025	Сбер, «Р-Фарм», AIRI	Соглашение /альянс	Discovery, preclinical	Внедрение передовых цифровых технологий в фарме; заявлены продукты и сервисы, включая R&D-направления и ускорение разработки	*https://pharmmedprom.ru/news/sber-r-farm-i-institut-airi-zaklyuchili-soglashenie-o-czifrovizaczii-farmaczevtiki/*
24.11.2025	«Р-Фарм», AIRI	Совместный R&D-проект	Discovery, preclinical	Применение генеративного ИИ для анализа заболеваний на молекулярном уровне и генерации/отбора молекул-кандидатов	*https://remedium.ru/news/r-farm-i-airi-pri-podderzhke-s/*

Из выходящих на рынок стартапов наиболее заметны команды, которые строят продукты вокруг цифровой химии и работы с химическими данными. Syntelly развивает платформу для извлечения и структурирования химической информации и поддержки задач органического синтеза и анализа химпространства, что делает ее близкой к прикладным задачам medicinal chemistry и ранней оптимизации кандидатов. На стыке вычислительной биологии и поиска новых терапевтических направлений выделяется Gero, позиционирующая себя как AI drug discovery компания, работающая с крупными массивами клинико-биологических данных для выявления мишеней и разработки терапевтиков, в том числе в области возраст-ассоциированных заболеваний. На практике это отражает наиболее реалистичный для России профиль стартапов в раннем R&D, не полный цикл от модели до клиники, а специализированные платформы, которые закрывают ключевые вычислительные этапы и могут подключаться к проектам фармкомпаний и научных организаций.

В клиническом контуре наиболее заметны стартапы, выросшие из диагностики и анализа медицинских данных и затем расширяющие применимость к задачам клинической разработки. OneCell (цифровая онкопатология) и UNIM (облачная цифровая патология) важны как источники стандартизируемых патоморфологических признаков и инфраструктуры для работы со слайдами, что напрямую связано с биомаркерами и оценкой ответа в онкологических программах. Celsus и Third Opinion развивают AI-платформы анализа медицинских изображений и внедряются на уровне крупных

клинических контуров, что повышает зрелость инфраструктуры для использования изображений как измеримых признаков в исследованиях. Webiomed и Botkin.AI фигурируют в профильных рейтингах российских мед-ИИ компаний и представляют класс решений, работающих с клиническими данными и поддержкой принятия решений; в контексте разработки лекарств такие инструменты наиболее релевантны в задачах стратификации, формирования когорт и повышения качества данных в клинических исследованиях (таблица 2.2).

Таблица 2.2 — Стартапы на российском рынке платформенной инфраструктуры

Стартап	Фокус	Контур	Описание	Источник
Syntelly	Хемоинформатика для задач органической и медицинской химии	Discovery, preclinical	Закрывает вычислительный слой «дизайн – отбор – оценка» свойств на ранних стадиях	*https://syntelly.com/*
Gero	AI-подходы для выявления биологии/мишеней и разработки терапевтиков	Discovery	Пример AI-drug discovery компании с фокусом на данных человека и таргетах	*https://www.gero.ai/*
OneCell	AI-цифровая патология и аналитика тканей	Clinical	Релевантно для биомаркеров и стандартизируемых признаков в онкопрограммах и клинических исследованиях	*https://www.onecell.ai/*
UNIM	End-to-end цифровая патология с нейросетевыми инструментами	Clinical	Инфраструктура для масштабируемой работы со слайдами и извлечения количественных признаков	*https://unim.su/digital_pathology*
Ligand Pro	AI-driven drug discovery и drug design	Discovery, preclinical	Развивает цифровые инструменты для генерации малых молекул, оценки взаимодействий «белок – лиганд», прогнозирования ADME-Tox и оптимизации белков	*https://ligandpro.ru/*
PeptiGen (проект компании «Альбоген»)	Drug design (пептидные молекулы)	Discovery, preclinical	ИИ-платформа для генерации и оптимизации пептидных соединений; публично подтверждено привлечение инвестиций в проект PeptiGen	*https://peptigen.com/; https://rb.ru/news/albogen-investments/*
EvoPreClinical	Цифровизация доклинических исследований	Preclinical	Платформа цифровизации доклинических исследований и молекул	*https://sechenov.tech/startups*

2.4. Барьеры роста и точки развития

Ключевой барьер роста российского рынка связан с данными и их промышленной пригодностью. Для задач discovery и preclinical критичны большие массивы воспроизводимых экспериментальных данных, но на практике они остаются фрагментированными по организациям и проектам, с неоднородными протоколами и ограниченной сопоставимостью, что снижает переносимость моделей и удлиняет цикл внедрения[85]. В клиническом контуре проблема еще острее: качество и полнота электронных медицинских данных ограничивают масштабируемую аналитику и обучение моделей, а высокая доля неупорядоченных записей требует значительных затрат на структурирование и контроль качества, что прямо отмечается в обсуждениях формирования рынка медицинских данных в России[86]. В нашей стране отсутствует нормативно закрепленный стандарт представления индивидуальных данных клинических исследований. В отличие от FDA и EMA, где обязательны форматы CDISC SDTM/ADaM (с 2016-го и 2023 года соответственно), российские досье содержат лишь агрегированные таблицы и PDF-отчеты. Это блокирует автоматизированную верификацию, метаанализ, развитие ИИ-инструментов и интеграцию с данными реальной клинической практики.

Второй крупный барьер относится к регуляторно-правовой среде и режимам доступа к информации. Медицинские данные и сведения, составляющие врачебную тайну, могут передаваться и использоваться в исследованиях при соблюдении строгих условий, включая согласие и иные основания, что повышает транзакционные издержки на получение данных, юридическую проработку и построение комплаенс-контуров[87]. На уровне отраслевой политики также фиксируется необходимость гармонизации регулирования разработки и регистрации лекарств с крупнейшими фармацевтическими рынками, поскольку различия в требованиях и процедурах могут ограничивать интеграцию российских разработок в международные цепочки R&D и усложнять масштабирование продуктов и подходов, ориентированных на глобальные стандарты доказательности.

Третий барьер — кадровый и организационный. Для устойчивого внедрения ИИ в разработку лекарств нужны междисциплинарные команды с компетенциями в химии, биологии, клинических исследованиях, data science, инженерии данных, а дефицит специалистов в фарминдустрии остается заметным и публично обозначается на уровне оценки потребности в кадрах до 2030 года. На практике это приводит к тому, что при наличии перспективных моделей и пилотов масштабирование упирается в способность организации выстроить производственный контур вокруг данных и моделей, поддерживать экспериментальную валидацию и формализовать процессы, чтобы решения могли тиражироваться между программами, а не оставались проектной активностью отдельных команд. Потребность фармацевтической отрасли в специалистах с высшим образованием до 2030 года превышает 7200 человек, текущая потребность — более 1500. Согласно анализу hh.ru, 68% вакансий с высшим образованием не закрыты, особенно по профилю «научный специалист, исследователь». Доля целевого приема по специальности «фармация» в 5–6 раз ниже, чем по «лечебному делу». Большинство молодых специалистов не видят целостной картины разработки, не понимают регуляторной логики и экономических ограничений[88].

Точки развития и условия ускорения в России в 2022–2026 гг. формируются вокруг кооперационных моделей и платформенных инициатив, которые снижают барьеры данных и компетенций. Один из векторов — стратегические альянсы и совместные программы крупных игроков, например Сбер, «Р-Фарм» и AIRI, а также Сбер, AIRI и ПРОМОМЕД, где заявлена цель развивать и внедрять AI-решения по цепочке создания лекарств, то есть фактически создавать общий технологический контур и распределять затраты на данные, вычисления и прикладную разработку. Второй вектор включает развитие отечественных платформ и пилотирование прикладных инструментов для химии и прогнозирования свойств, что снижает порог входа для R&D-команд и формирует базовый инструментальный слой рынка. Устойчивый рост происходит там, где одновременно развиваются инфраструктура данных, включая повышение качества и структурированности клинической информации, комплаенс-процессы и кооперационные форматы, позволяющие наращивать датасеты и воспроизводимо переносить модели между программами и организациями.

2.5. Правовое регулирование в России

В России правовое регулирование применения ИИ в фармацевтической отрасли пока не оформлено в виде самостоятельной и детализированной системы, сопоставимой с профильными международными подходами, однако соответствующий контур постепенно складывается на пересечении государственной политики в сфере ИИ, регулирования обращения лекарственных препаратов, требований к медицинским данным и норм, действующих в смежных сегментах цифрового здравоохранения. На практике это означает, что развитие ИИ в фарме в России происходит в условиях частично фрагментированной нормативной среды, где сохраняется потребность в более четкой регламентации требований к качеству данных, прозрачности алгоритмов, доказательству корректности их применения и порядку использования таких решений в исследованиях, разработке и регуляторных процедурах. В документах РФ и ЕАЭС на данный момент отсутствуют рекомендации или руководства по использованию математического моделирования и ИИ в разработке лекарств. В отличие от ICH M15, национальная нормативная база не содержит специализированных требований к валидации и документированию AI/ML-моделей на этапах drug discovery и клинических исследований[89] [90] [91].

Указ Президента РФ № 490 «О развитии искусственного интеллекта в Российской Федерации», утвердивший Национальную стратегию развития ИИ до 2030 года, задает для российского рынка базовую технологическую и институциональную рамку, в пределах которой развивается применение ИИ в том числе в фармацевтической отрасли. Его значение для рассматриваемого сегмента заключается не в установлении специальных требований к разработке лекарственных препаратов, а в формировании общего приоритета внедрения ИИ, развитии инфраструктуры данных, повышении доверия к цифровым решениям и стимулировании практического использования таких технологий в экономике и социальной сфере. Для фармацевтического рынка этот документ важен как стратегический источник, определяющий направление движения отрасли, но не заменяющий профильное регулирование исследований, регистрации и обращения препаратов[92].

Федеральный закон № 323-ФЗ «Об основах охраны здоровья граждан в Российской Федерации» формирует общий правовой контур функционирования системы здравоохранения, в который неизбежно встраиваются и цифровые решения, применяемые на стыке фармы, клинической практики и медицинских данных. Практическая роль этого закона в теме ИИ связана с тем, что именно он закрепляет базовые принципы охраны здоровья, права пациента, требования к медицинской информации и общие условия осуществления медицинской деятельности, что делает его опорным актом для любых сценариев, где фармацевтическая разработка соприкасается с клиническими данными, реальной медицинской практикой и цифровыми инструментами сопровождения лечения. В результате закон выступает не специализированным актом об ИИ, а фундаментом, через который определяется допустимость и пределы интеграции таких решений в медицинский контур[93].

Федеральный закон № 61-ФЗ «Об обращении лекарственных средств» является ключевым отраслевым актом для всей цепочки создания и вывода препарата на рынок, включая разработку, доклинические и клинические исследования, регистрацию, производство, контроль качества и обращение лекарственных средств. В контексте ИИ его значение определяется тем, что в рамках закона проходит основной регуляторный маршрут, в который потенциально встраиваются алгоритмические инструменты, используемые для отбора молекул, анализа данных исследований, оптимизации дизайна разработок и подготовки регуляторных материалов. При этом сам закон пока не содержит развернутого специального режима для ИИ в фармацевтическом R&D, что создает ситуацию, при которой внедрение таких решений происходит внутри общего контура лекарственного регулирования, но без детализированной отраслевой настройки именно под алгоритмические модели[94]. В ЕАЭС действует процедура ускоренной регистрации (100 рабочих дней вместо 140) для препаратов особой значимости, орфанных, детских и снижающих смертность. Однако адаптивные дизайны клинических исследований прямо не регламентированы, хотя обновленные правила GCP ЕАЭС поощряют рискориентированные подходы. Руководства по биостатистике (Рекомендация ЕЭК № 19) основаны на ICH, но руководства по ИИ/ММ отсутствуют[95].

Кодекс этики применения ИИ в сфере здравоохранения не является нормативным актом прямого действия, однако выполняет важную функцию мягкого регулирования, закрепляя набор принципов, на которые ориентируются участники рынка при внедрении ИИ в медицинские процессы. В опубликованных материалах Минздрава подчеркивается, что кодекс включает следующие принципы: ограниченная автономность, прозрачность, конфиденциальность данных, справедливость, этическая экспертиза и контроль, — то есть формирует рамку доверенного и социально приемлемого использования технологий. Для фармацевтической отрасли этот документ значим прежде всего как этический ориентир в тех сегментах, где ИИ соприкасается с медицинскими данными, клиническими решениями и пациентской безопасностью, хотя сам по себе он не устанавливает обязательных процедур для регистрации или разработки лекарственных препаратов[96].

ГОСТ Р 71671-2024 «Системы поддержки принятия врачебных решений с применением искусственного интеллекта. Основные положения» представляет собой уже более предметный элемент российского регулирования, поскольку

переводит обсуждение ИИ из уровня общих принципов в плоскость стандартизации конкретного класса медицинских систем. По данным Росстандарта, документ действует с 1 января 2025 года и относится к системам поддержки принятия врачебных решений с применением ИИ, что указывает на формирование в России унифицированного подхода к описанию и структурированию таких решений в клинической среде. Для фармацевтического рынка значение этого ГОСТа также носит преимущественно смежный характер и относится прежде всего к медицинским информационным и клиническим системам, показывает, что российская регуляторная среда постепенно переходит к более детализированной стандартизации ИИ-решений в здравоохранении, включая требования к базовой архитектуре и назначению подобных систем[97].

Приказ Росздравнадзора от 21.07.2025 № 4472 формирует один из наиболее прикладных и формализованных элементов российского контура регулирования ИИ в медицине, поскольку утверждает порядок автоматической передачи в информационную систему Росздравнадзора сведений об обрабатываемых данных и результатах действия программного обеспечения с применением технологий ИИ, если такое программное обеспечение является медицинским изделием и обращается на территории Российской Федерации. Практическое значение приказа заключается в том, что регулятор переходит от общего надзора к более системному сбору данных о функционировании ИИ-систем в реальном обращении, что усиливает требования к прослеживаемости, мониторингу и пострегистрационному контролю таких решений. Для анализа рынка ИИ в фарме этот документ важен как индикатор институциональной зрелости смежного сегмента: он по-прежнему относится к программному обеспечению как медицинскому изделию, а не к собственно алгоритмам, применяемым в разработке лекарств, однако демонстрирует движение российского регулирования в сторону постоянного наблюдения за результатами работы ИИ-систем после их вывода в практическое использование[98].

Таким образом, правовое регулирование применения ИИ в российской фармацевтической отрасли находится на этапе формирования и пока развивается преимущественно через сочетание общих стратегических, отраслевых, этических и смежных медицинских норм, а не через единый специализированный режим, ориентированный именно на AI-drug discovery и цифровое сопровождение разработки лекарственных препаратов. Это означает, что базовые элементы регулирования уже присутствуют от уровня государственной политики в сфере ИИ, законодательств о здравоохранении и лекарственных средствах до требований к работе с данными и механизмов надзора за отдельными классами ИИ-систем. Однако сама нормативная среда пока остается фрагментированной и в значительной степени опирается на адаптацию существующих правовых конструкций к новым технологическим сценариям.

На рынке такая ситуация создает двойственный эффект. С одной стороны, наличие базовой правовой и институциональной рамки позволяет развивать пилоты, формировать кооперации и постепенно накапливать практику применения ИИ в фарме и смежных сегментах здравоохранения. С другой стороны, отсутствие детализированных специальных требований к качеству моделей, допустимым сценариям использования, валидации,

документированию и регуляторному статусу алгоритмических решений в контуре разработки лекарств сдерживает масштабирование и повышает неопределенность для участников рынка. В перспективе дальнейшее развитие будет зависеть от того, насколько последовательно российская нормативная среда сможет перейти от общего регулирования цифровых технологий к более прикладному и отраслево-специфическому контуру, сопоставимому с формирующимися международными подходами.

В ответ на описанные регуляторные пробелы представителями индустрии в составе ХелсНет НТИ, Р-Фарм, MS Decisions и Институтом искусственного интеллекта разработано Рамочное соглашение о принципах использования ИИ в жизненном цикле лекарственных средств (полный текст см. в Приложении 2). Документ предлагает риск-ориентированную модель регулирования, основанную на принципах пропорциональности, качества данных, прозрачности и непрерывного мониторинга, и призван стать основой для разработки национальных руководств по применению ИИ в фармацевтическом R&D.

2.6. Первоочередные шаги по формированию национальных гайдлайнов

Принимая во внимание отсутствие детализированных регуляторных требований к применению ИИ в фармацевтической разработке и одновременно наличие консолидированной отраслевой позиции, зафиксированной в Рамочном соглашении (Приложение 2), первостепенной задачей становится формирование системы методических руководств и индустриальных стандартов. Это позволит дать участникам рынка предсказуемые ориентиры и накопить практический опыт, необходимый для последующей формализации требований в нормативных правовых актах (НПА).

Ниже представлен план первоочередных действий, основанный на принципах «мягкого регулирования» и инициативах «снизу-вверх». План не требует немедленных изменений федерального законодательства и может быть реализован силами отраслевого сообщества при координации с профильными ведомствами.

Шаг 1. Разработка и утверждение методических рекомендаций по валидации AI/ML-моделей в контексте лекарственной разработки. В течение 12 месяцев с момента подписания Рамочного соглашения отраслевой консорциум (фармкомпании, институты развития, профильные ассоциации, ФОИВы и другие присоединившиеся участники) готовит проект методических рекомендаций, определяющих:

- минимальные требования к планированию и документированию вычислительного эксперимента (план анализа модели);
- критерии приемлемости качества данных, адаптированные для задач фармы;
- стандартную структуру отчёта о валидации, включая анализ устойчивости и переносимости модели;
- подходы к управлению дрейфом модели и процедуры повторной оценки.

Подготовленный проект передаётся на рассмотрение в Минздрав России и/или подведомственные экспертные организации в рамках процедуры научного консультирования. После получения обратной связи документ утверждается на уровне отраслевого объединения и рекомендуется к применению при формировании досье и взаимодействии с регуляторными органами.

Шаг 2. Пилотирование подходов в формате «регуляторных песочниц». Для апробации разрабатываемых гайдлайнов и накопления прецедентов инициировать 2–3 пилотных проекта (например, виртуальный скрининг с последующей экспериментальной валидацией или использование цифровых конечных точек в клиническом исследовании), в рамках которых модели сопровождаются по согласованной с регулятором программе валидации. По итогам пилотов формируются публичные кейсы, демонстрирующие, каким образом применение ИИ повышает качество и эффективность разработки без ущерба для безопасности пациента. Взаимодействие с регулятором на данном этапе может осуществляться в формате отдельных научных консультаций или через механизмы, предусмотренные действующим законодательством об обращении лекарственных средств.

Шаг 3. Разработка национального стандарта (ГОСТ Р). Параллельно с накоплением опыта инициировать в Росстандарте разработку ГОСТ Р «Системы искусственного интеллекта в фармацевтической разработке. Общие требования», опираясь на международные наработки (ISO/IEC 22989, ISO/IEC 23053, NIST AI RMF, ICH M15) и принципы, зафиксированные в Приложении 2. Наличие ГОСТа создаст единую терминологическую и процессную рамку, а главное — станет мостом к будущему включению таких требований в официальные регуляторные процедуры. Координацию работ по стандартизации целесообразно вести через профильные технические комитеты (ТК 164 «Искусственный интеллект», ТК 459 «Информационные технологии в здравоохранении») с привлечением экспертов консорциума.

Шаг 4. Создание открытых бенчмарков и отраслевых референсных датасетов. Одним из главных ограничений остаётся дефицит стандартизированных наборов данных для объективного сравнения моделей. Целесообразно в рамках консорциума инициировать создание защищённого депозитария анонимизированных экспериментальных и клинических данных (с учётом законодательства о персональных данных и интеллектуальной собственности) и разработать регламент их использования для целей бенчмаркинга AI-моделей. Это не требует изменения НПА, но нуждается в подписании многостороннего соглашения о партнёрстве между участниками консорциума.

Шаг 5. Образовательные модули и повышение квалификации. Совместно с профильными вузами и Евразийской академией надлежащих практик запустить модули по AI для сотрудников регуляторных органов и R&D-персонала компаний. Цель — формирование общего языка и понимания методологии валидации моделей, без чего невозможна ни эффективная экспертиза, ни индустриальное масштабирование.

Шаг 6. Внедрение стандартизированных форматов клинических данных (CDISC). Инициировать переход к обязательному использованию форматов CDISC SDTM/ADaM при формировании регистрационных досье в ЕАЭС и РФ (по

аналогии с FDA и EMA). Разработать дорожную карту на 2–3 года с пилотными проектами. Это снимет фундаментальное ограничение для автоматизированной верификации, мета-анализа и обучения ИИ-моделей на сопоставимых данных.

Реализация перечисленных шагов не требует немедленной корректировки федерального законодательства и может быть начата силами отраслевого сообщества при координации с профильными ведомствами. По мере накопления достаточного объёма прецедентов и формирования устойчивой методической базы станет возможным и следующий этап: формализация апробированных требований в виде поправок в подзаконные акты или специализированных разделов регистрационных руководств ЕАЭС.

2.7. Практики применения ИИ в российских проектах 2022–2025 гг.

В 2022–2025 гг. в России сформировался массив научных проектов, в которых ИИ и смежные вычислительные методы применяются в контуре фармацевтических исследований как инструмент ускорения и повышения точности ранних этапов разработки. Анализ этих практик показывает, что российский рынок движется преимущественно по модели точечного внедрения ИИ в отдельные исследовательские задачи от поиска и отбора перспективных соединений до анализа биологической активности и поддержки экспериментальной валидации, полноценные сквозные платформенные решения пока распространены меньше. Проекты представлены в Приложении.

В проектах 2022–2025 гг. ИИ и смежные вычислительные методы применяются прежде всего для решения задач раннего этапа фармацевтической разработки, где требуется ускорить поиск и первичный отбор перспективных соединений. Наиболее частотными сценариями выступают виртуальный скрининг, QSAR-моделирование, молекулярный докинг, прогноз биологической активности и поиск новых фармакологических свойств уже известных веществ. Такая логика прослеживается, в частности, в проекте «Создание новых лекарственных веществ с помощью искусственного интеллекта» ФГАОУ ВО «РУДН им. Патриса Лумумбы», где in silico-подходы используются для прогнозирования биологической активности и молекулярных механизмов связывания производных левофлоксацина, а также в проекте по разработке методологии компьютерного поиска фармакологически активных соединений на основе множественного докинга и искусственных нейронных сетей ФГБОУ ВО ВолгГМУ Минздрава России, где нейросетевой прогноз применяется для отбора веществ с заданной психотропной и RAGE-ингибирующей активностью. Сходный подход реализован и в проекте по виртуальному скринингу лигандов для липидных GPCR-рецепторов, где МО и высокопроизводительные вычисления используются как инструмент рационального дизайна лекарств на основе структуры белка-мишени.

Отдельное направление российских практик связано с использованием ИИ и вычислительной аналитики для репозиционирования препаратов, анализа больших массивов данных и поддержки выбора терапевтических гипотез в условиях инфекционных угроз. Это особенно заметно в проектах «Репозиционирование лекарств для терапии COVID-19» ИБМХ и «Анализ больших мультидисциплинарных данных для оптимизации создания и применения

фармакотерапевтических средств в условиях коронавирусных эпидемий» РФФИ, где вычислительные платформы, прогнозирование in silico и методы МО применяются для отбора кандидатов, оценки межлекарственных взаимодействий и информационно-аналитической поддержки поиска антикоронавирусных решений. При этом российская практика в целом носит гибридный характер, ИИ, как правило, не замещает экспериментальный этап, а используется для сужения воронки поиска с последующей проверкой результатов in vitro и in vivo. Это видно и в проекте по изучению противовирусного действия новых азолоазиновых и азолопиримидиновых молекул, где виртуальный скрининг сочетается с молекулярной динамикой и экспериментальной оценкой связывания, и в проекте по созданию нового класса веществ с антидепрессивной и противовоспалительной активностями, где вычислительный drug design дополняется синтезом и биологическим тестированием. Тем самым российские проекты 2022–2025 гг. показывают, что ИИ уже встроен в прикладной исследовательский контур фармы, однако его основная роль пока сосредоточена на ускорении поиска, приоритизации и интерпретации данных, а не на формировании полноценных сквозных платформ разработки препарата.

Массив проектов также показывает, что практики применения ИИ в фарме развиваются не только вокруг поиска отдельных молекул, но и вокруг постепенного формирования собственных вычислительных методологий. Это особенно заметно в проектах, где исследовательская цель состоит не просто в подборе соединения, а в разработке самого инструментария отбора и анализа. К таким примерам относится проект по созданию методологии компьютерного поиска фармакологически активных соединений на основе множественного докинга и искусственных нейронных сетей, где акцент сделан на повышении точности прогноза психотропной и RAGE-ингибирующей активности, а также на построении воспроизводимого вычислительного контура для дальнейшего поиска кандидатов. Наличие подобных проектов позволяет говорить о том, что часть российских инициатив постепенно выходит за пределы разовых прикладных расчетов и движется в сторону создания собственных исследовательских платформ и методик, пригодных для многократного использования в лекарственной разработке, хотя и остается на начальном уровне.

Проекты показывают, что российский контур остается в значительной степени научно-исследовательским и концентрируется вокруг университетов, научных институтов и грантовых механизмов, а не вокруг масштабных индустриальных платформ полного цикла. По содержанию работ видно, что даже наиболее продвинутые проекты ориентированы прежде всего на решение узких научных задач, таких как поиск лигандов для конкретных мишеней, моделирование противовирусной активности, изучение молекулярных механизмов связывания, формирование наборов данных и последующую точечную экспериментальную проверку. Такая структура подтверждает: практика применения ИИ пока развивается преимущественно как сеть отдельных исследовательских треков, где вычислительные методы используются для повышения эффективности раннего этапа разработки, но еще не формируют широко распространенную модель сквозного платформенного сопровождения препарата от гипотезы до поздних стадий вывода на рынок.

ЧАСТЬ III
БИБЛИОГРАФИЧЕСКИЙ ПОИСК

3. Библиографический поиск

3.1. Методология исследования

Библиографический поиск по базе данных портфолио журналов Nature за 2020–2026 гг. с ключевыми словами: Artificial intelligence in drug development, Drug discovery, Drug design, Preclinical development, Clinical development, Target identification and validation, ADMET, Toxicity prediction, Cheminformatics, QSAR / QSPR показал 461 результат[99].

Запрос формировался следующим образом: («Artificial intelligence in drug development» OR «Artificial intelligence») AND («Drug discovery» OR «Drug design») AND («Preclinical development» OR «Clinical development» OR «Target identification and validation» OR «ADMET» OR «Toxicity prediction» OR «Cheminformatics» OR «QSAR» OR «QSPR»).

Логика запроса следующая: мы разделили ключевые слова на три логических блока, объединенных оператором AND. Это означает, что в результатах поиска обязательно должны присутствовать элементы из всех трех смысловых групп:

1. Технология (AI): мы ищем либо точную фразу «Artificial intelligence in drug development», либо просто «Artificial intelligence», чтобы не пропустить статьи, где формулировка отличается.
2. Область применения (General Scope): используются «Drug discovery» или «Drug design». Статья должна касаться открытия или дизайна лекарств.
3. Специфика и методы (Specific Methodologies & Stages): перечислены все узкоспециализированные термины через оператор OR. Это заставит поиск найти статьи, которые содержат хотя бы один из этих терминов (ADMET, QSAR, Toxicity prediction и т. д.) в контексте первых двух групп.

3.2. Приоритеты в научных исследованиях за последние пять лет

На основе цитируемых и значимых статей мы составили хронологию исследовательских приоритетов. Из списка ниже видно четкое развитие от скрининга (2020) к структурной биологии (2021) и наконец к генеративным моделям и мультимодальным взаимодействиям (2023–2025).

2024–2025: Эра фундаментальных моделей (Foundation Models) и предсказания взаимодействий

Приоритет исследований: в 2024 году и с переходом в 2025-й фокус сместился с предсказания структуры *одиночного* белка на предсказание взаимодействий всех типов биомолекул (белок – белок, белок – ДНК, белок – лиганд). Приоритетом стало создание «фундаментальных моделей» (по аналогии с GPT), которые понимают биологический контекст целиком. Главное событие — выход AlphaFold 3, который моделирует не только форму, но и химическое взаимодействие лекарства с мишенью с беспрецедентной точностью

Статья (определяющая тренд 2024/25): Abramson J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3 // Nature. — 2024. — Vol. 630. — pp. 493–500. Ссылка: *https://www.nature.com/articles/s41586-024-07487-w*

2023: Взрыв генеративного дизайна и диффузионных моделей (Diffusion Models)

Приоритет исследований: этот год ознаменовался переходом от *аналитического* ИИ (который классифицирует молекулы) к *генеративному* (который их создает). В drug design стали использоваться диффузионные модели (технология, лежащая в основе генерации изображений), адаптированные для создания белков *de novo*. Исследователи научились создавать белковые структуры, которых не существует в природе, под конкретные задачи.

Статья: Watson J. L. et al. De novo design of protein structure and function with RFdiffusion // Nature. — 2023. — Vol. 620. — pp. 1089–1100. Ссылка: *https://www.nature.com/articles/s41586-023-06415-8.*

2022: Освоение химического пространства сверхвысокой размерности (Chemical Space Exploration)

Приоритет исследований: в 2022 году приоритет был направлен на масштабирование. Использование ИИ позволило работать с библиотеками «синтонов», содержащими десятки миллиардов соединений. Приоритетом стало не просто найти лекарство, а найти его в максимально доступном химическом пространстве, используя графовые алгоритмы для поиска редких активных молекул, которые легко синтезировать.

Статья: Sadybekov A. A. et al. Synthon-based ligand discovery in virtual libraries of billions of compounds // Nature. — 2022. — Vol. 601. — pp. 452–459. Ссылка: *https://www.nature.com/articles/s41586-021-04220-9.*

2021: Решение проблемы фолдинга белка (The AlphaFold Moment)

Приоритет исследований: 2021 год вошел в историю как переломный момент для *Target identification*. Публикация AlphaFold2 решила 50-летнюю проблему биологии. Приоритетом всех исследований стало массовое применение структурного предсказания для всего протеома человека. Это дало старт эпохе Structural-based drug design без необходимости проведения дорогостоящей рентгеноструктурной кристаллографии для каждого белка.

Статья: Jumper J. et al. Highly accurate protein structure prediction with AlphaFold // Nature. — 2021. — Vol. 596. — pp. 583–589. Ссылка: *https://www.nature.com/articles/s41586-021-03819-2.*

2020: Демократизация виртуального скрининга и борьба с резистентностью

Приоритет исследований: в 2020 году на фоне пандемии и роста интереса к вычислительным методам приоритетом стало создание открытых платформ для *Virtual Screening*. Исследования фокусировались на том, чтобы сделать скрининг миллионов молекул доступным для любой лаборатории (Open Source), а также на применении Deep Learning для поиска новых классов антибиотиков, отличных от известных (уход от резистентности).

Статья: Gorgulla C. et al. An open-source drug discovery platform enables ultra-large virtual screens // Nature. — 2020. — Vol. 580. — pp. 663–668. Ссылка: *https://www.nature.com/articles/s41586-020-2117-z.*

3.3. Тренды в публикациях за 2022–2026 гг.

Для глубокого аналитического обзора мы использовали журналы из портфолио *Nature* (включая *Nature*, *Nature Medicine*, *Nature Biotechnology*, *Nature Machine Intelligence*, *Nature Reviews Drug Discovery*).

Ниже представлены тренды, обозначенные в публикации научных статей из портфолио Nature по использованию ИИ (AI) в drug discovery / drug design / preclinical & clinical development / target ID&validation / ADMET / toxicity / cheminformatics / QSAR/QSPR.

1. Deep QSAR / Deep QSPR: интеграция QSAR/QSPR с deep learning и рост «предсказательного дизайна»

Суть тренда: QSAR/QSPR переходит от «классических дескрипторов + регрессия» к глубоким моделям (graph/transformer/NN), комбинируясь с генеративным дизайном, планированием синтеза и виртуальным скринингом. Параллельно растет прикладная QSPR-аналитика (антибиотики, НПВС, антималярийные ряды) с ML.

Статьи:

1) **Huang Y. et al.** On QSPR analysis for predicting efficacy of physicochemical properties of antibiotics drugs via topological indices and regression models // Journal of Molecular Graphics and Modelling. — 2026. — Vol. 144. — Art. No. 109280. Ссылка: *https://www.sciencedirect.com/science/article/abs/pii/S1093326326000069.*
2) **Ahmed W. et al.** Computational approaches in drug chemistry leveraging python powered QSPR study of antimalaria compounds by using artificial neural networks // Scientific Reports. — 2025. — Vol. 15. — Art. No. 19307. Ссылка: *https://www.nature.com/articles/s41598-025-01594-y.*
3) **Hashem A. F. et al.** TOPSIS based multi criteria QSPR modeling of antibiotics using graph theoretic indices // Scientific Reports. — 2025. — Vol. 15. — Art. No. 40658. Ссылка: *https://www.nature.com/articles/s41598-025-25565-5.*
4) **Tropsha A. et al.** Integrating QSAR modelling and deep learning in drug discovery: the emergence of deep QSAR // Nature Reviews Drug Discovery. — 2024. — Vol. 23. — pp. 141–155. Ссылка: *https://www.nature.com/articles/s41573-023-00832-0.*

2. ADMET/Toxicity: сдвиг «left» (раньше) + explainable AI и переносимость/реюз моделей

Суть тренда: токсикология и ADMET становятся «ранними фильтрами» в дизайне. Усиливается фокус на интерпретируемости (explainable AI), переносимости моделей между задачами и наборами данных и на прикладных моделях для органоспецифической токсичности и метаболизма (например, CYP450).

Статьи:

1) **Amin, S.A., Kar, S. & Piotto, S.** KidneyTox_v1.0 enables explainable artificial intelligence prediction of nephrotoxicity in small molecules. *Sci Rep* 16, 5099 (2026). Ссылка: *https://doi.org/10.1038/s41598-026-35496-4.*
2) **Atwereboannah, A.A., Wu, WP., Al-antari, M.A.** ***et al.*** MEN: leveraging explainable multimodal encoding network for precision prediction of CYP450

inhibitors. *Sci Rep* 15, 21820 (2025). Ссылка: *https://doi.org/10.1038/s41598-025-04982-6.*

3) **Li, R., Lu, J., Liu, Z. *et al.*** Reusability report: exploring the utility of variational graph encoders for predicting molecular toxicity in drug design. *Nat Mach Intell* 6, 1457–1466 (2024). Ссылка: *https://doi.org/10.1038/s42256-024-00923-6.*

3. Data-centric AI: качество данных, курация, бенчмаркинг «под реальный мир»

Суть тренда: приоритет смещается с «какая модель лучше» к «какие данные и как измеряем». Появляются специализированные бенчмарки и схемы сплита для устранения смещений; курация датасетов становится ключевым фактором прироста качества; растет роль data governance в организациях open science.

Статьи:

1) **Tian, T., Li, S., Zhang, Z. et al.** Benchmarking compound activity prediction for real-world drug discovery applications. *Commun Chem* 7, 127 (2024). Ссылка: *https://doi.org/10.1038/s42004-024-01204-4.*
2) **Edfeldt, K., Edwards, A.M., Engkvist, O. et al.** A data science roadmap for open science organizations engaged in early-stage drug discovery. *Nat Commun* 15, 5640 (2024). Ссылка: *https://doi.org/10.1038/s41467-024-49777-x.*
3) **Meng, J., Chen, P., Wahib, M. et al.** Boosting the predictive performance with aqueous solubility dataset curation. *Sci Data* 9, 71 (2022). Ссылка: *https://doi.org/10.1038/s41597-022-01154-3.*

4. Operationalization: «конвейеры» AI/ML для виртуального скрининга и ускорения discovery (в том числе в ресурсоограниченных средах)

Суть тренда: AI перестает быть разрозненными моделями и оформляется в воспроизводимые end-to-end-пайплайны (screening cascade, практики внедрения), включая сценарии реального внедрения в центрах с ограниченной инфраструктурой.

Статьи:

1) **Yu, S., Peng, Q., Wei, W. et al.** AI-driven virtual screening platform identifies novel NSUN2 inhibitor candidates for targeted cancer therapy: a computational drug discovery approach. *npj Precis. Onc.* 10, 98 (2026). Ссылка: *https://doi.org/10.1038/s41698-026-01296-2.*
2) **Turon, G., Hlozek, J., Woodland, J.G. et al.** First fully-automated AI/ML virtual screening cascade implemented at a drug discovery centre in Africa. *Nat Commun* 14, 5736 (2023). Ссылка: *https://doi.org/10.1038/s41467-023-41512-2.*
3) **Sadybekov, A.V., Katritch, V.** Computational approaches streamlining drug discovery. *Nature* 616, 673–685 (2023). Ссылка: *https://doi.org/10.1038/s41586-023-05905-z.*

5. AI для lead optimization и multi-parameter optimization: от фрагментов и карманов белка до «предпочтений химика»

Суть тренда: модели все чаще оптимизируют не один параметр (affinity), а набор целей (potency/селективность/свойства/синтетичность). Важный пласт — обучение на человеческих предпочтениях (medchem intuition) и pocket-aware-оптимизация.

Статьи:

1) **Jeon, H., Lee, J.G., Shin, W. et al.** STELLA provides a drug design framework enabling extensive fragment-level chemical space exploration and balanced multi-parameter optimization. *Sci Rep* 15, 28135 (2025). Ссылка: *https://doi.org/10.1038/s41598-025-12685-1.*
2) **Chen, S., Zhang, O., Jiang, C. et al.** Deep lead optimization enveloped in protein pocket and its application in designing potent and selective ligands targeting LTK protein. *Nat Mach Intell* 7, 448–458 (2025). Ссылка: *https://doi.org/10.1038/s42256-025-00997-w.*
3) **Choung, OH., Vianello, R., Segler, M. et al.** Extracting medicinal chemistry intuition via preference machine learning. *Nat Commun* 14, 6651 (2023). Ссылка: *https://doi.org/10.1038/s41467-023-42242-1.*

6. Новые представления молекул: LLM/токенизация, self-supervised и «визуальные» представления + перенос на свойства/мишени

Суть тренда: идет борьба за «правильное представление» молекулы (tokens/граф/3D/изображение), что напрямую улучшает property prediction, conformation generation и pocket-based generation. Набирают силу self-supervised и (квази-)мультимодальные подходы.

Статьи:

1) **Wang, J., Qin, R., Wang, M. et al.** Token-Mol 1.0: tokenized drug design with large language models. *Nat Commun* 16, 4416 (2025). Ссылка: *https://doi.org/10.1038/s41467-025-59628-y.*
2) **Wang, S., Zhang, R., Li, X. et al.** Recent advances in molecular representation methods and their applications in scaffold hopping. *npj Drug Discov.* 2, 14 (2025). Ссылка: *https://doi.org/10.1038/s44386-025-00017-2.*
3) **Zeng, X., Xiang, H., Yu, L. et al.** Accurate prediction of molecular properties and drug targets using a self-supervised image representation learning framework. *Nat Mach Intell* 4, 1004–1016 (2022). Ссылка: *https://doi.org/10.1038/s42256-022-00557-6.*

7. Precision oncology: от target identification к клинической трансляции (включая ADC и мультитаргет)

Суть тренда: в онкологии AI используется как «сквозной ускоритель» — поддержка выбора мишеней, дизайн молекул/конъюгатов и переход к клинической значимости (персонализация, стратификация, мультитаргет-стратегии).

Статьи:

1) **Zhou, J., Jiang, Y., Yu, M. et al.** Kinic index: an artificial intelligence-driven predictive model and multitarget drug discovery framework for hepatocellular carcinoma patients. *NPJ Precis. Onc.* (2026). Ссылка: *https://doi.org/10.1038/s41698-026-01324-1.*
2) **Sutanto, H., Fetarayani, D.** Integrating artificial intelligence into small molecule development for precision cancer immunomodulation therapy. *npj Drug Discov.* 2, 25 (2025). Ссылка: *https://doi.org/10.1038/s44386-025-00029-y.*
3) **Lu, Y., Huang, W., Li, Y. et al.** Leveraging artificial intelligence in antibody-drug conjugate development: from target identification to clinical translation in

oncology. *npj Precis. Onc.* 9, 374 (2025). Ссылка: *https://doi.org/10.1038/s41698-025-01159-2.*

8. Выход за пределы «малых молекул»: large molecules, RNAi, PROTAC + рынок и инфраструктура

Суть тренда: AI становится инструментом развития новых модальностей (large-molecule discovery, RNAi design, PROTAC), а также анализа индустриального ландшафта и данных (патентные датасеты), которые питают модели.

Статьи:

1) **Cai, H., Yao, G., Shi, Y. et al.** PROTAC-PatentDB: A PROTAC Patent Compound Dataset. *Sci Data* 12, 1840 (2025). Ссылка: *https://doi.org/10.1038/s41597-025-06136-9.*
2) **Tang, Q., Khvorova, A.** RNAi-based drug design: considerations and future directions. *Nat Rev Drug Discov* 23, 341–364 (2024). Ссылка: *https://doi.org/10.1038/s41573-024-00912-9.*
3) **Nagra N. S. et al.** The company landscape for artificial intelligence in large-molecule drug discovery. *Nat Rev Drug Discov* 22, 949–950 (2023) Ссылка: *https://doi.org/10.1038/d41573-023-00139-0.*

9. Инфекционные заболевания и природные соединения: AI как ускоритель поиска противовирусных/антибактериальных кандидатов и natural products

Суть тренда: AI все активнее применяется там, где особенно важны скорость и масштаб (антивирусные программы, антибиотики, natural products). Одновременно подчеркиваются барьеры: сложность биологии, данные, перевод в разработку.

Статьи:

1) **Du, S., Hu, X., Li, P. et al.** Antiviral drug discovery and development: challenges and future directions. *Sig Transduct Target Ther* 11, 69 (2026). Ссылка: *https://doi.org/10.1038/s41392-025-02539-7.*
2) **Mullowney, M.W., Duncan, K.R., Elsayed, S.S. et al.** Artificial intelligence for natural product drug discovery. *Nat Rev Drug Discov* 22, 895–916 (2023). Ссылка: *https://doi.org/10.1038/s41573-023-00774-7.*
3) **Miethke, M., Pieroni, M., Weber, T. et al.** Towards the sustainable discovery and development of new antibiotics. *Nat Rev Chem* 5, 726–749 (2021). Ссылка: *https://doi.org/10.1038/s41570-021-00313-1*.

3.4. Анализ трендов в методологических подходах в исследованиях использования моделей ИИ в драгдизайне

Ниже представлены 7 ключевых трендов с подборкой из трех-пяти наиболее часто цитируемых статей на каждый.

Тренд 1. Структурная революция: AlphaFold и предсказание взаимодействия с мишенью

Анализ: технологии глубокого обучения решили проблему «фолдинга» белков. Тренд сместился от простого предсказания формы белка к

предсказанию того, как малые молекулы или другие белки связываются с этими структурами, открывая возможности для дизайна лекарств под ранее недоступные мишени.

Статьи:

1) **Jumper J. et al.** Highly accurate protein structure prediction with AlphaFold // Nature. — 2021. — Vol. 596. — pp. 583–589. Ссылка: *https://www.nature.com/articles/s41586-021-03819-2.*
2) **Tunyasuvunakool K. et al.** Highly accurate protein structure prediction for the human proteome // Nature. — 2021. — Vol. 596. — pp. 590–596. Ссылка: *https://www.nature.com/articles/s41586-021-03828-1.*
3) **Ren F. et al.** AlphaFold accelerates artificial intelligence powered drug discovery: efficient discovery of a novel CDK20 small molecule inhibitor // Chemical science. — 2023. — Vol. 14. — No. 6. — pp. 1443–1452. Ссылка: *https://pubs.rsc.org/en/content/articlehtml/2023/sc/d2sc05709c.*
4) **Callaway E.** 'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures // Nature. — 2020. — Vol. 588. — pp. 203–204. Ссылка: *https://www.nature.com/articles/d41586-020-03348-4.*

Тренд 2. Генеративный ИИ для дизайна de novo (Generative Chemistry)

Анализ: использование генеративных состязательных сетей (GANs), вариационных автокодировщиков (VAE) и диффузионных моделей. ИИ не просто ищет в базе, а создает новые химические сущности, оптимизированные под мультипараметрические цели (активность + синтезируемость).

Статьи:

1) **Du Y. et al.** Machine learning-aided generative molecular design // Nature Machine Intelligence. — 2024. — Vol. 6. — No. 6. — pp. 589–604. Ссылка: *https://www.nature.com/articles/s42256-024-00843-5.*
2) **Flam-Shepherd D. et al.** Language models can learn complex molecular distributions // Nature Communications. — 2022. — Vol. 13. — Art. No. 3293. Ссылка: *https://www.nature.com/articles/s41467-022-30839-x.*
3) **Tong X. et al.** Generative models for de novo drug design // Journal of Medicinal Chemistry. — 2021. — Vol. 64. — No. 19. — pp. 14011–14027. Ссылка: *https://pubs.acs.org/doi/abs/10.1021/acs.jmedchem.1c00927.*
4) **Griffiths R. R., Hernández-Lobato J. M.** Constrained Bayesian optimization for automatic chemical design using variational autoencoders // Chemical science. — 2020. — Vol. 11. — No. 2. — pp. 577–586. Ссылка: *https://pubs.rsc.org/en/content/articlehtml/2019/sc/c9sc04026a.*

Тренд 3. Масштабируемый виртуальный скрининг (Giga-scale Screening)

Анализ: взрывной рост библиотек для скрининга (до миллиардов молекул) благодаря новым алгоритмам докинга и облачным вычислениям. Это позволяет находить «иголки в стоге сена», которые раньше пропускались из-за ограниченности библиотек.

Статьи:

1) **Sadybekov A. A. et al.** Synthon-based ligand discovery in virtual libraries of billions of compounds // Nature. — 2022. — Vol. 601. — pp. 452–459. Ссылка: *https://www.nature.com/articles/s41586-021-04220-9.*

2) **Gorgulla C. et al.** An open-source drug discovery platform enables ultra-large virtual screens // Nature. — 2020. — Vol. 580. — pp. 663–668. Ссылка: *https://www.nature.com/articles/s41586-020-2117-z.*
3) **Lyu J. et al.** Ultra-large library docking for discovering new chemotypes // Nature. — 2019. — Vol. 566. — pp. 224–229. Ссылка: *https://www.nature.com/articles/s41586-019-0917-9.*

Тренд 4. Автоматизация синтеза и автономные лаборатории (Self-driving Labs)

Анализ: интеграция ИИ с робототехникой. Алгоритмы планируют синтез (ретросинтез), а роботы его исполняют. ИИ анализирует результат и корректирует план следующего эксперимента без участия человека.

Статьи:

1) **Burger B. et al.** A mobile robotic chemist // Nature. — 2020. — Vol. 583. — pp. 237–241. Ссылка: *https://www.nature.com/articles/s41586-020-2442-2.*
2) **Schwaller P. et al.** Molecular transformer: a model for uncertainty-calibrated chemical reaction prediction // ACS central science. — 2019. — Vol. 5. — No. 9. — pp. 1572–1583. Ссылка: *https://pubs.acs.org/doi/full/10.1021/acscentsci.9b00576.*
3) **Coley C. W. et al.** A robotic platform for flow synthesis of organic compounds informed by AI planning // Science. — 2019. — Vol. 365. — No. 6453. — Art. No. eaax1566. Ссылка: *https://www.science.org/doi/full/10.1126/science.aax1566.*
4) **Segler M. H. S. et al.** Planning chemical syntheses with deep neural networks and symbolic AI // Nature. — 2018. — Vol. 555. — pp. 604–610 (классика, задавшая тренд). Ссылка: *https://www.nature.com/articles/nature25978.*

Тренд 5. ИИ в клинических испытаниях и Precision Medicine

Анализ: использование ИИ для анализа данных реального мира (RWE) и омиксных данных для отбора пациентов. Цель — снизить процент неудач на 2–3 фазах испытаний за счет точной стратификации больных.

Статьи:

1) **Topol E. J.** High-performance medicine: the convergence of human and artificial intelligence // Nature Medicine. — 2019. — Vol. 25. — No. 1. — pp. 44–56. Ссылка: *https://www.nature.com/articles/s41591-018-0300-7.*
2) **Harrer S. et al.** Artificial intelligence for clinical trial design // Trends in pharmacological sciences. — 2019. — Vol. 40. — No. 8. — pp. 577–591. Ссылка: *https://www.cell.com/trends/pharmacological-sciences/fulltext/S0165-6147(19)30130-0?dgcid=raven_jbs_etoc_email.*
3) **Vamathevan J. et al.** Applications of machine learning in drug discovery and development // Nature Reviews Drug Discovery. — 2019. — Vol. 18. — pp. 463–477. Ссылка: *https://www.nature.com/articles/s41573-019-0024-5.*

Тренд 6. Прогнозирование ADMET и безопасности (Safety First)

Анализ: переход от «проверки в конце» к «прогнозу в начале». Использование мультизадачного обучения (multitask learning) для одновременного предсказания токсичности, растворимости и метаболической стабильности.

Статьи:

1) **Zhang K. et al.** Artificial intelligence in drug development // Nature medicine. — 2025. — Vol. 31. — No. 1. — pp. 45–59. Ссылка: *https://www.nature.com/articles/s41591-024-03434-4.*
2) **Das S. K. et al.** Deep learning: A game changer in drug design and development // Advances in Pharmacology. — Academic Press, 2025. — Vol. 103. — pp. 101–120. Ссылка: *https://www.sciencedirect.com/science/chapter/bookseries/abs/pii/S1054358925000080.*
3) **Pathan I. et al.** Revolutionizing pharmacology: AI-powered approaches in molecular modeling and ADMET prediction // Medicine in Drug Discovery. — 2025. — Art. No. 100223. Ссылка: *https://www.sciencedirect.com/science/article/pii/S259009862500020X.*
4) **Catacutan D. B. et al.** Machine learning in preclinical drug discovery // Nature Chemical Biology. — 2024. — Vol. 20. — No. 8. — pp. 960–973. Ссылка: *https://www.nature.com/articles/s41589-024-01679-1.*
5) **Fang X. et al.** Geometry-enhanced molecular representation learning for property prediction // Nature Machine Intelligence. — 2022. — Vol. 4. — No. 2. — pp. 127–134. Ссылка: *https://www.nature.com/articles/s42256-021-00438-4.*
6) **Bender A., Cortés-Ciriano I.** Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 1: Ways to make an impact, and why we are not there yet // Drug discovery today. — 2021. — Vol. 26. — No. 2. — pp. 511–524. Ссылка: *https://www.sciencedirect.com/science/article/pii/S1359644620305274.*

Тренд 7. Интеграция мультиомики и системная фармакология

Анализ: отход от принципа «одно лекарство — одна мишень» к системному подходу. ИИ анализирует графы знаний (Knowledge Graphs), объединяя геномику, протеомику и клинические данные для поиска неочевидных связей и перепрофилирования лекарств.

Статьи:

1) **Li L. et al.** Kolmogorov–Arnold graph neural networks for molecular property prediction // Nature Machine Intelligence. — 2025. — Vol. 7. — No. 8. — pp. 1346–1354. Ссылка: *https://www.nature.com/articles/s42256-025-01087-7.*
2) **Zhang O. et al.** Graph neural networks in modern AI-aided drug discovery // Chemical Reviews. — 2025. — Vol. 125. — No. 20. — pp. 10001–10103. Ссылка: *https://pubs.acs.org/doi/abs/10.1021/acs.chemrev.5c00461.*
3) **Corso G. et al.** Graph neural networks // Nature Reviews Methods Primers. — 2024. — Vol. 4. — Art. No. 17. Ссылка: *https://www.nature.com/articles/s43586-024-00294-7.*
4) **Bonner S. et al.** A review of biomedical datasets relating to drug discovery: a knowledge graph perspective // Briefings in Bioinformatics. — 2022. — Vol. 23. — No. 6. — Art. No. bbac404. Ссылка: *https://academic.oup.com/bib/article/23/6/bbac404/6712301?login=false&guestAccessKey.*
5) **Li M. M. et al.** Graph representation learning in biomedicine and healthcare // Nature Biomedical Engineering. — 2022. — Vol. 6. — pp. 1353–1369. Ссылка: *https://www.nature.com/articles/s41551-022-00942-x.*

6) **Schneider P. et al.** Rethinking drug design in the artificial intelligence era // Nature Reviews Drug Discovery. — 2020. — Vol. 19. — pp. 353–364. Ссылка: *https://www.nature.com/articles/s41573-019-0050-3.*

МНЕНИЯ ЭКСПЕРТОВ

В рамках отчета были опрошены ряд экспертов, профессионально имеющих отношение к использованию прорывных цифровых технологий в фармакологии и при разработке новых лекарственных препаратов, специалистов, владеющих информацией по использованию искусственного интеллекта в программных продуктах и прикладных разработках.

Экспертам предлагалось развернуто ответить на два вопроса, связанных с описанием трендов в области развития искусственного интеллекта в драгдизайне и описанием российских перспективных проектов и компаний.

Эксперты сходятся во мнении, что искусственный интеллект уже стал важным инструментом современной разработки лекарственных препаратов, однако его роль существенно различается на разных этапах этого процесса. Наиболее заметные результаты ИИ демонстрирует в поиске и валидации биологических мишеней, прогнозировании структуры белков, виртуальном скрининге и дизайне малых молекул, что позволяет значительно ускорить исследования и сократить объем экспериментальной работы. Вместе с тем ИИ не рассматривается как замена лабораторных и клинических испытаний: его основная задача заключается в повышении эффективности принятия решений, снижении числа неудачных гипотез и оптимизации процесса разработки при сохранении требований к доказательности, безопасности и этическим стандартам.

В качестве ключевых направлений дальнейшего развития эксперты выделяют переход от отдельных алгоритмов к комплексным цифровым исследовательским платформам, объединяющим генеративный ИИ, фундаментальные модели для биологии и химии, математическое моделирование, цифровых двойников и автоматизацию исследовательского цикла. Особое значение приобретает интеграция данных различных типов — от молекулярной информации до клинических данных — для более точного прогнозирования эффективности и безопасности препаратов на всех этапах разработки. При этом наряду с развитием технологий возрастают требования к защите данных, воспроизводимости результатов и соответствию регуляторным требованиям, что делает информационную безопасность и доверие к ИИ-платформам одним из ключевых факторов их практического применения.

Кирилл Песков

руководитель Центра математического моделирования в разработке лекарств ПМГУ им. Сеченова

Искусственный интеллект в разработке лекарственных препаратов: от дизайна молекул к клинической фармакометрике

Прежде всего, необходимо провести четкую границу между понятиями «дизайн препарата» (drug design), «поиск препарата» (drug discovery) и «разработка препарата» (drug development). В отечественном информационном поле, ввиду исторического дефицита проектов по созданию оригинальных молекул, эти термины часто смешиваются, хотя они обозначают принципиально разные этапы. Искусственный интеллект играет на этих этапах совершенно разные роли. Если говорить о дизайне малых молекул, то использование ИИ, глубокого (Deep Learning) и машинного обучения (Machine Learning) уже давно не является инновационным трендом или так называемым «game changer» — это устоявшийся отраслевой стандарт. Глобальные фармацевтические компании активно применяют эти инструменты как минимум последние 10–15 лет, в первую очередь на этапах идентификации и оптимизации лидерных соединений (lead identification и lead optimization). Квалифицированный разработчик сегодня по умолчанию использует эти алгоритмы. Однако существует критический нюанс, неочевидный для широкой аудитории: успешное применение ИИ на ранней стадии химического синтеза и дизайна малых молекул практически никак не коррелирует с вероятностью успеха на поздних фазах клинических исследований. В этом заключается уникальный разрыв, свойственный современной фармакологии.

Наиболее значимый прогресс последнего десятилетия, который можно классифицировать как историю успеха, наблюдается на этапе drug discovery, а именно — в поиске и валидации биологических мишеней. По сути, это область фундаментальных поисковых исследований. Долгое время наука не имела эффективных инструментов для перехода от аминокислотной последовательности к пониманию пространственной структуры белка без проведения сложных и зачастую безрезультатных эмпирических экспериментов. Сегодня появление таких нейросетевых инструментов, как AlphaFold, позволяет с высокой точностью предсказывать 3D-структуру макромолекул. Это дает исследователям первичное представление о терапевтическом потенциале мишени — так называемой «драггабильности» (druggability). Получив структуру, мы можем определить наличие доменов, с которыми способны взаимодействовать субстраты или лиганды, и понять, имеет ли смысл разрабатывать малую молекулу для таргетирования, ингибирования или активации данного белка. Это позволяет значительно снизить долю работы «вслепую».

Существует мнение, что развитие больших языковых моделей (LLM) и генеративного ИИ позволит в будущем полностью миновать лабораторные стадии in vitro/in vivo и сразу переходить к клинической практике. Однако к подобным заявлениям следует относиться с крайней осторожностью. Часто это

не более чем агрессивный PR со стороны IT-компаний, заходящих в сферу биологии. В разработке лекарственных препаратов существуют строгие этические и научные стандарты. Выход в клинику (тестирование на людях) без предварительного формирования надежной доказательной базы на животных моделях абсолютно неэтичен и недопустим, равно как и регистрация препарата без доказанной эффективности и безопасности.

Цель применения ИИ — не замена экспериментальной биологии или клинических испытаний, а их рационализация. Строгий инженерный подход и математическое моделирование позволяют отсеять бессмысленный массив данных и сузить пространство эмпирического поиска до минимально необходимого количества ключевых исследований. ИИ помогает быстрее сгенерировать нужную структуру с заданными фармакологическими свойствами, убедиться в принципиальной возможности ее химического синтеза и проверить на патентную чистоту. Все это существенно сокращает время вывода препарата на рынок (time-to-market), но не отменяет необходимости самих испытаний.

Все описанные выше алгоритмы показывают высокую эффективность именно при работе с малыми молекулами. Когда речь заходит о крупных биопрепаратах (например, моноклональных антителах), вес методов искусственного интеллекта в их дизайне пока значительно ниже.

Главная проблема заключается в том, что при дизайне молекулы исследователи зачастую оптимизируют лишь один-два параметра — например, константу связывания с мишенью (аффинность). Однако высокая аффинность ничего не говорит о фармакокинетике препарата: как он циркулирует в организме, как распределяется по тканям и метаболизируется (так называемые drug-like properties). Например, перспективный онкологический препарат должен эффективно проникать в опухолевую ткань и накапливаться в ней, но при этом минимизировать проникновение в костный мозг, где также активно делятся, но уже здоровые клетки. Если предсказать этот фактор только на этапе дизайна не удастся, в клинике препарат продемонстрирует неприемлемый профиль токсичности и будет снят с исследований. Именно поэтому ключевой тренд и потребность завтрашнего дня — это интеграция хемоинформатики (drug design) с методами фармакометрического моделирования. Речь идет о создании комплексной среды, связывающей генерацию структур с физиологически обоснованным фармакокинетическим моделированием (PBPK) и количественной системной фармакологией (QSP).

В идеале мы должны получить «in silico архитектурное бюро»: ИИ генерирует молекулу, определяет ее базовые фармакологические параметры, а математическая модель физиологически обоснованно предсказывает концентрацию препарата в конкретных органах (эффект-компартментах) и прогнозирует фармакодинамический ответ у конкретной популяции пациентов. Подобная линковка методов позволит нивелировать текущий разрыв между успешным начальным дизайном и провалами на поздних стадиях трансляционной медицины. Интерес к этому направлению огромен: недавно крупный инвестиционный фонд приобрел западную компанию Simulation Plus, специализирующуюся именно на фармакометрическом моделировании, за сумму свыше 300 млн долларов.

На отечественном рынке присутствуют конкурентоспособные команды. В основном они функционируют в формате сервисных компаний,

специализирующихся на дизайне малых молекул с применением алгоритмов машинного обучения (например, проекты ООО «Синтелли», ООО «Лиганд Про», а также специалисты, ранее работавшие в Insilico Medicine). Они предоставляют грамотные услуги по генерации молекул по запросу. Однако комплексные платформы, объединяющие драг-дизайн с клинической фармакометрикой (по аналогии с тем, что мы сейчас развиваем на базе компании SimurgFarm), — это более сложная и уникальная задача. Именно за созданием интегрированных систем, способных предсказывать судьбу препарата от химического синтеза до ответа организма пациента, стоит будущее фармацевтической разработки.

Пермяков Руслан Анатольевич
заместитель директора ЦК НТИ ТДВ,
доверенный эксперт ХелсНет,
эксперт SafeNet

Развитие ИИ в области drug deign сегодня переходит от этапа использования точечных решений к формированию полноценных цифровых исследовательских платформ по принципы AI-scientist. Такие платформы объединяют автономных агентов обеспечивающих автоматизацию от сбора данны, и проверки гипотез в разных областях биологии, вычислительной химии и т.п. до автоматизации лабораторных исследований. Наиболее важные изменения происходят не столько в совершенствовании отдельных моделей машинного обучения, сколько в изменении всей архитектуры процесса Drug Discovery.

Ключевые мировые тренды

Первый и наиболее заметный тренд — переход от специализированных моделей к универсальным foundation-моделям для молекулярной биологии и химии. Например, пару лет назад ИИ решал отдельные задачи (QSAR, поиск мишеней, виртуальный скрининг), а сегодня появляются решения, одновременно работающие с белковыми последовательностями, химическими структурами, трехмерной геометрией молекул, научными публикациями и экспериментальными данными.

Второй тренд — интеграция различных источников информации для повышения качества прогнозов и моделей. Сейчас ведутся исследования учитывающие в прогнозах геномные данные, химические библиотеки, изображения клеток, электронные медицинские карты и научную литературу, данные фитнес-браслетов, в плоть до покупок в магазине.

Отдельного внимания заслуживает быстрое развитие технологий цифровых двойников. Сегодня речь идет уже не только о цифровых двойниках молекул, но и о цифровых двойниках клеток, тканей, пациентов и отдельных стадий разработки лекарственного препарата. В перспективе именно цифровые двойники способны объединить результаты моделирования, лабораторных экспериментов и клинических исследований в единую непрерывную модель жизненного цикла препарата.

С моей точки зрения, важным вызовом становится обеспечение безопасности и приватности при использовании искусственного интеллекта в таких исследованиях. Современные модели обучаются на огромных массивах научных и клинических данных, значительная часть которых содержит коммерческую тайну, результаты еще не опубликованных исследований или персональные медицинские данные. Это требует внедрения комплексных механизмов защиты на всех этапах жизненного цикла ИИ — от безопасного хранения и обезличивания данных до контроля доступа, защиты моделей от различных атак, обеспечения воспроизводимости результатов и аудита принимаемых решений. Думаю, в ближайшие годы конкурентоспособность ИИ-платформ будет определяться не только качеством алгоритмов, но и их способностью обеспечивать высокий уровень информационной безопасности, соответствие регуляторным требованиям и доверие со стороны исследователей, бизнеса и государства.

Перспективные проекты и технологии есть в России наиболее важные и интересные

Я бы здесь отметил следующие коллективы:

- AIRI — фундаментальные исследования ИИ для биологии и медицины;
- Сколтех — применение ИИ в вычислительной биологии и дизайне белков;
- Институт биоорганической химии РАН — структурная биология и белковая инженерия;
- Институт цитологии и генетики СО РАН — геномика и биоинформатика;
- BIOCAD — одна из наиболее активно использующих ИИ российских биофармацевтических компаний;
- Сбер AI и Yandex Research — большие модели, которые постепенно адаптируются для биомедицинских задач;
- проекты в области генеративной химии (Molecule Generation, QSAR), развиваемые университетскими и академическими группами.

Анна Пинигина
ведущий хемоинформатик
ООО «Синтелли»

Ключевые тренды в области развития ИИ в драгдизайне

Ключевых трендов в развитии ИИ в драгдизайне несколько. Прежде всего это генеративный дизайн — нейросети, которые конструируют новые молекулы и белки с нуля под заданные свойства, позволяя перебирать огромные виртуальные пространства соединений вместо того, чтобы синтезировать тысячи веществ вручную.

Отдельным прорывом стал вычислительный дизайн белков, который был отмечен Нобелевской премией по химии 2024 года. Система AlphaFold и появившееся вслед за ней семейство открытых моделей сделали предсказание пространственной структуры белков и их комплексов с лекарствами доступными широкому кругу исследователей. Недавно к этому добавилась оценка силы связывания молекулы с мишенью.

Развиваются также ИИ-докинг, модели токсикологии и фармакокинетики, а на переднем крае — автономные лаборатории и агентные системы, которые замыкают цикл проектирования, синтеза и анализа. Важно отметить, что такие лаборатории ускоряют работу, но не заменяют исследователя, оставаясь инструментом в руках эксперта.

Перспективные проекты и технологии есть в России наиболее важные и интересные

Россия представлена несколькими сильными, хотя и разрозненными командами. Среди индустриальных игроков выделяется BIOCAD с развитым направлением вычислительной биологии и собственными подходами к разработке антител. Группа компаний «ХимРар» строит платформу полного цикла — от поиска мишени до клинических этапов. Есть и международные проекты с российскими корнями, такие как Insilico Medicine.

Важную импортозаместительную роль играют и технологические платформы нового поколения. Здесь не могу обойти вниманием Синтелли. Она формирует единую инфраструктуру для работы с химическими данными и ИИ-моделями, и выступает отечественным аналогом зарубежных решений в области химических баз данных и инструментов анализа.

Активно развиваются и академические центры: ИТМО с ассистентом химика и мультиагентными системами генерации молекул, AIRI совместно со Сбером, создавшие центр разработки лекарств с применением ИИ и линейку биологических моделей, а также Сколтех, МФТИ и институты РАН.

Государственная поддержка реализуется через стратегию развития фармацевтической промышленности, национальные программы по ИИ и технологическому суверенитету.

Российская экосистема пока уступает мировым лидерам, однако формирует фундамент для ускоренного роста за счет сочетания сильной научной школы и развивающихся технологических решений.

Заключение

Проведённое исследование подтверждает, что внедрение искусственного интеллекта в фармакологию и разработку лекарственных средств перешло из стадии экспериментальных пилотов в фазу промышленной трансформации. Глобальный рынок демонстрирует устойчивый рост, консолидацию вокруг платформенных решений и усиление требований к качеству данных, валидации моделей и регуляторной определённости.

На этом фоне российский сектор находится на этапе активного формирования: завершён период точечных научных проектов, и отрасль вплотную подошла к необходимости создания национальной платформенной архитектуры, согласованной с международными стандартами, но учитывающей локальную специфику. Ключевыми барьерами остаются фрагментация данных, отсутствие нормативно закреплённых требований к ИИ-моделям в R&D и кадровый дефицит. В то же время кооперационные проекты (Сбер, Р-Фарм, AIRI, ПРОМОМЕД) и разработанное Рамочное соглашение о принципах использования ИИ в жизненном цикле лекарственных средств закладывают основу для системных изменений.

Представленный в отчёте план первоочередных действий (валидационные методики, регуляторные песочницы, ГОСТ Р, отраслевые бенчмарки, образовательные модули, внедрение CDISC) формирует дорожную карту на ближайшие 2–3 года. Его реализация позволит России не только сократить отставание от лидеров, но и занять конкурентные позиции в формирующемся глобальном ландшафте ИИ-ускоренной фармацевтической разработки.

Авторский коллектив намерен продолжать системный мониторинг развития темы, отслеживая изменения в нормативной среде, технологические прорывы и появление новых отечественных решений. В следующих отраслевых отчётах мы планируем оценить прогресс в реализации обозначенных шагов, проанализировать emerging-тренды и представить новые российские кейсы. Приглашаем всех участников рынка к сотрудничеству и диалогу — только совместными усилиями можно выстроить устойчивую экосистему ИИ в фармацевтической отрасли России.

Список источников

[1] https://www.esmorwd.org/article/S2949-8201%2826%2900001-9/fulltext
[2] https://www.fda.gov/about-fda/center-drug-evaluation-and-research-cder/artificial-intelligence-drug-development
[3] https://www.fda.gov/about-fda/artificial-intelligence-drug-development/guiding-principles-good-ai-practice-drug-development
[4] https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle_en.pdf
[5] https://www.annualreviews.org/content/journals/10.1146/annurev-pharmtox-040323-040828;jsessionid=OSGczETRsnuoc3zH5NY6QI4z28yfv0oRS7h1jTkc.annurevlive-10-241-10-93
[6] https://www.sciencedirect.com/science/article/pii/S1359644625002314
[7] https://www.science.org/doi/10.1126/science.adk8946
[8] https://academic.oup.com/nsr/article/13/1/nwaf495/8320534
[9] https://www.nature.com/articles/s41746-025-01992-6
[10] https://alphafoldserver.com/welcome
[11] https://www.nature.com/articles/s41586-024-07487-w
[12] https://royalsocietypublishing.org/rsos/article/12/7/250646/235354/Autonomous-self-driving-laboratories-a-review-of
[13] https://www.nature.com/articles/s41597-024-03793-0
[14] https://www.nature.com/articles/s41467-025-61355-3
[15] https://www.sciencedirect.com/science/article/pii/S1386505625003582
[16] https://chemrxiv.org/doi/full/10.26434/chemrxiv-2022-jjm0j-v4
[17] https://pubs.acs.org/doi/full/10.1021/jacsau.4c01160
[18] https://www.nature.com/articles/s41467-026-69302-6
[19] https://arxiv.org/pdf/2401.03369
[20] https://academic.oup.com/bib/article/27/1/bbaf703/8418201
[21] https://pmc.ncbi.nlm.nih.gov/articles/PMC11383977
[22]https://www.sciencedirect.com/science/chapter/bookseries/abs/pii/S0065774325000028
[23] https://www.nature.com/articles/s44385-024-00003-9
[24] https://pubs.acs.org/doi/10.1021/acs.chemrev.5c00461
[25] https://pubs.rsc.org/en/content/articlehtml/2025/sc/d4sc07875f
[26] https://www.nature.com/articles/s41586-023-06415-8
[27] https://www.nature.com/articles/s41586-025-09721-5
[28] https://academic.oup.com/bioinformatics/article/40/7/btae416/7698030
[29] https://tdcommons.ai/benchmark/admet_group/overview/
[30] https://pubs.acs.org/doi/10.1021/acs.jcim.4c02122
[31] https://ai.jmir.org/2024/1/e50800/
[32] https://www.nature.com/articles/s41467-024-53081-z
[33] https://pmc.ncbi.nlm.nih.gov/articles/PMC12476210
[34] https://pmc.ncbi.nlm.nih.gov/articles/PMC12861029
[35] https://portlandpress.com/emergtoplifesci/article/8/2/89/236530/The-impact-of-artificial-intelligence-on-biomarker
[36] https://www.astrazeneca.com/media-centre/press-releases/2025/astrazeneca-enters-into-collaboration-with-cspc.html

[37] https://www.pfizer.com/about/partners/research-and-business-development-partnerships
[38] https://www.astrazeneca.com/what-science-can-do/topics/data-science-ai/mapdiff-edge-set-attention-ai-protein-design-breakthroughs.html
[39] https://www.gene.com/stories/redefining-drug-discovery-with-ai
[40] https://www.roche.com/investors/updates/inv-update-2023-11-21
[41] https://www.roche.com/media/releases/med-cor-2025-04-29
[42] https://insilico.com/pipeline
[43] https://insilico.com/news/u4cbsok2s1-insilico-medicine-announces-us888-millio
[44] https://www.schrodinger.com/platform/products/livedesign/
[45] https://en.xtalpi.com/
[46] https://www.prnewswire.com/news-releases/xtalpi-and-pfizer-expand-strategic-collaboration-to-advance-ai-driven-drug-discovery-and-materials-science-simulations-302494033.html
[47] https://www.isomorphiclabs.com/articles/isomorphic-labs-kicks-off-2024-with-two-pharmaceutical-collaborations
[48] https://www.isomorphiclabs.com/articles/isomorphic-labs-announces-novartis-collaboration-expansion
[49] https://www.isomorphiclabs.com/partnerships
[50] https://pharmaphorum.com/news/jj-bets-isomorphic-ai-powered-drug-hunt
[51] https://www.iqvia.com/newsroom/2025/06/iqvia-launches-new-ai-agents-for-life-sciences-and-healthcare
[52] https://www.iqvia.com/newsroom/2025/01/iqvia-and-nvidia-collaborate-to-transform-healthcare-and-life-sciences
[53] https://www.tempus.com/news/pr/tempus-expands-ai-enabled-care-pathway-intelligence-platform-into-breast-cancer/
[54] https://events.flatiron.com/asco2025
[55] https://aws.amazon.com/ru/health/life-sciences-symposium-2025/
[56] https://www.arizton.com/market-reports/ai-in-drug-discovery-market
[57] https://www.gminsights.com/industry-analysis/ai-in-drug-discovery-market
[58] https://www.researchandmarkets.com/reports/5595828/artificial-intelligence-in-drug-discovery-market
[59] https://www.sciencedirect.com/science/article/pii/S2590098625000107
[60] https://pmc.ncbi.nlm.nih.gov/articles/PMC11444559
[61] https://www.marketsandmarkets.com/Market-Reports/ai-in-clinical-trials-market-42687548.html
[62] https://www.mordorintelligence.com/industry-reports/ai-in-clinical-trials-market
[63] https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-based-clinical-trials-solution-provider-market
[64]https://database.ich.org/sites/default/files/ICH_E6%28R3%29_Step4_FinalGuideline_2025_0106.pdf
[65] https://www.fda.gov/regulatory-information/search-fda-guidance-documents/conducting-clinical-trials-decentralized-elements
[66] https://pharmvestnik.ru/content/news/minpromtorg-rf-ocenil-deficit-kadrov-dlya-farmproma-do-2030-goda.html
[67] https://mosapteki.ru/material/obrazovatelnye-standarty-podgotovki-farmrabotnikov-budut-peresmotreny-31347
[68] https://www.ich.org/news/harmonised-ich-m15-guideline-general-principles-model-informed-drug-development-adopted

[69] https://www.ich.org/page/articles-procedures
[70] https://pubmed.ncbi.nlm.nih.gov/41874058/
[71]https://database.ich.org/sites/default/files/ICH_Step_2_Presentation_M15_2024_1129.pdf
[72] https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-artificial-intelligence-support-regulatory-decision-making-drug-and-biological
[73] https://www.ispor.org/heor-resources/presentations-database/presentation-cti/ispor-2026/poster-session-3-3/breakthrough-therapy-bt-and-fast-track-ft-designations-in-new-drug-applications-and-biologic-license-applications-ndas-blas-to-fda-s-center-for-drug-evaluation-and-research-cder--2013-2025
[74] https://www.federalregister.gov/documents/2025/01/07/2024-31542/considerations-for-the-use-of-artificial-intelligence-to-support-regulatory-decision-making-for-drug
[75] https://pharmreviews.kz/novosti/zarubeznye-novosti/voz-opredelila-clinical-trials-information-system-osnovnym-registrom-klinicheskikh-ispytanij
[76] https://www.ema.europa.eu/en/use-artificial-intelligence-medicinal-product-lifecycle
[77]https://www.sberbank.ru/ru/sberpress/all/article?blockID=1303&lang=ru&newsID=a038e13b-69b2-41e8-9fb5-7b1ef105277d&regionID=77&type=NEWS
[78] https://sk.ru/news/novyy-rezident-skolkovo-zaymetsya-uvelicheniem-prodolzhitelnosti-zhizni-cheloveka/
[79] https://www.kommersant.ru/doc/7923399
[80] https://www.sberbank.ru/ru/sberpress/obrazovanie/article?blockID=cae5dec3-0369-4139-950d-56308a727059&lang=ru&newsID=e2b34059-f21d-4639-b426-82164018cb11&regionID=27&type=NEWS
[81] https://promomed.ru/news/sber-institut-airi-i-promomed-obedinilis-dlya-sozdaniya-novykh-lekarstv-s-pomoshchyu-ii
[82] https://highload.ru/spb/2022/abstracts/9112
[83] https://simurgpharm.ru/
[84]https://www.sberbank.ru/ru/sberpress/all/article?blockID=1303&lang=ru&newsID=a038e13b-69b2-41e8-9fb5-7b1ef105277d&regionID=77&type=NEWS
[85] https://static-0.minzdrav.gov.ru/system/attachments/attaches/000/057/382/original/%D0%A1%D1%82%D1%80%D0%B0%D1%82%D0%B5%D0%B3%D0%B8%D1%8F_%D1%86%D0%B8%D1%84%D1%80%D0%BE%D0%B2%D0%BE%D0%B9_%D1%82%D1%80%D0%B0%D0%BD%D1%81%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%86%D0%B8%D0%B8_%D0%BE%D1%82%D1%80%D0%B0%D1%81%D0%BB%D0%B8_%D0%97%D0%B4%D1%80%D0%B0%D0%B2%D0%BE%D0%BE%D1%85%D1%80%D0%B0%D0%BD%D0%B5%D0%BD%D0%B8%D0%B5.pdf
[86] https://medvestnik.ru/content/articles/vladimir-voskoboev-iskusstvennyi-intellekt-mojet-sdelat-medicinu-bolee-chelovechnoi.html
[87]https://www.consultant.ru/document/cons_doc_LAW_121895/9f906d460f9454a8a0d290738d9fc2798c1e865a
[88] https://gxpnews.net/2026/01/minpromtorg-obnarodoval-prognoz-po-kadrovomu-deficzitu-v-farmprome/
[89] https://www.lidings.com/ru/media/legalupdates/AI_pharma/
[90] https://www.regmed.ru/news/iskusstvennyy-intellekt-v-razrabotke-lekarstvennykh-sredstv/

[91] https://www.regmed.ru/ads/6-2025-zhurnala-regulyatornye-issledovaniya-i-ekspertiza-lekarstvennykh-sredstv/
[92] https://publication.pravo.gov.ru/document/view/0001201910110003
[93] https://publication.pravo.gov.ru/Document/View/0001201111220007
[94] https://publication.pravo.gov.ru/file/pdf?eoNumber=0001202210200012
[95] https://remedium.ru/news/eek-sformulirovala-rekomendats/
[96] https://minzdrav.gov.ru/news/2016/05/30/2974-minzdravom-rossii-razrabotan-proekt-kodeksa-professionalnoy-etiki-rabotnikov-sistemy-zdravoohraneniya-rossiyskoy-federatsii
[97] https://protect.gost.ru/document.aspx?control=7&id=263101
[98] https://zn1.roszdravnadzor.gov.ru/documents/87798
[99]https://www.nature.com/search?q=%28%22Artificial+intelligence+in+drug+development%22+OR+%22Artificial+intelligence%22%29+AND+%28%22Drug+discovery%22+OR+%22Drug+design%22%29+AND+%28%22Preclinical+development%22+OR+%22Clinical+development%22+OR+%22Target+identification+and+validation%22+OR+%22ADMET%22+OR+%22Toxicity+prediction%22+OR+%22Cheminformatics%22+OR+%22QSAR%22+OR+%22QSPR%22%29&date_range=last_5_years&order=relevance

Приложение 1

Проекты российских научных и образовательных организаций по изучению использования ИИ в драгдизайне

Наименование проекта	Описание проекта	Автор проекта	Этап разработки лекарств
Создание новых лекарственных веществ с помощью ИИ	Заключается в QSAR-изучении подходов in silico для прогнозирования биологической активности и молекулярных механизмов связывания производных левофлоксацина. В результате исследования на примере реальной (RMS) и предсказанной (PMS) структур левофлоксацина (Lvf) продемонстрировано использование подходов in silico для поиска количественных корреляций «структура – активность». Профиль биологической активности производных левофлоксацина оценивали в порядке убывания сложности их структуры по сравнению с исходным левофлоксацином на основе байесовской вероятностной модели. Показано, что отсутствие карбоксильной группы в структуре молекулы значительно снижает прогностическую антимикробную активность фторхинолонов (FlQns) в 2 раза. Кроме того, эти процессы приводят к появлению других «неклассических» биологических активностей, таких как ингибирование изоформы цитохрома P-450 CYP1A2. Показано, что наиболее информативными и слабо вырожденными топологическими индексами для описания и различения RMS- и PMS-производных Lvf являются индексы Балабана и Электропи. Судя по представленным 2D-диаграммам, структурами, близко похожими друг на друга, являются основной и дефторированный Lvf. С помощью программных продуктов нового поколения для молекулярного докинга (AutoDockTools 1.5.7 с программой нового поколения Autodock Vina) биологически активные молекулы-кандидаты левофлоксацина, отобранные по результатам QSAR-анализа, были визуализированы в виде 3D-моделей лиганд-рецепторных комплексов. На основании эмпирической скоринговой функции, конечной предсказанной энергии связывания (аффинности, ккал/моль), константы связывания Kb (M-	ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «РОССИЙСКИЙ УНИВЕРСИТЕТ ДРУЖБЫ НАРОДОВ ИМЕНИ ПАТРИСА ЛУМУМБЫ»	Поиск и оптимизация кандидатов на ранней стадии разработки

Наименование проекта	Описание проекта	Автор проекта	Этап разработки лекарств
	1), а также данных диаграмм кластеризации продемонстрирована близость структуры 9-дефторированного Lvf к исходной структуре левофлоксацина. Полученные результаты имеют практическое значение для внедрения в фармацию препаратов с заранее заданными свойствами, полученными на основе известного препарата-лидера и определенного активного центра белка-мишени		
Механизмы противовирусного действия новых лекарственных / перспективных молекул азолоазинового и азолопиримидинового ряда на in silico виртуальных и молекулярных и субклеточных моделях in vitro и in vivo	Отчет состоит из 53 страниц, содержит 23 рисунка, 4 схемы, 10 таблиц, 37 использованных источников. В результате выполнения этого этапа были наработаны и исследованы новые перспективные соединения, изучены антирадикальные свойства и их взаимодействие с различными ферментативными системами. Проведены исследования взаимодействия молекул с модельными системами in silico, проанализированы механизмы их антивирусной активности, получены данные по молекулярно-динамическим расчетам, оценены параметры связывания с белками вируса. Исследования позволили выявить перспективные молекулы для дальнейших этапов разработки противовирусных средств	ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «УРАЛЬСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ ИМЕНИ ПЕРВОГО ПРЕЗИДЕНТА РОССИИ Б.Н. ЕЛЬЦИНА»	Доклиническая разработка и валидация противовирусных кандидатов
Разработка нового класса веществ с антидепрессивной и противовоспалительной активностями. Drug design, синтез, биологическая активность	Объектом исследования является новый класс мультитаргетных соединений с выраженной противовоспалительной и антидепрессивной активностью. Цель работы — разработка подходов к дизайну, синтезу и изучению биологической активности новых соединений с данным типом фармакологического действия. В процессе выполнения проекта были предложены новые структуры веществ, проведен их синтез, выполнена экспериментальная оценка фармакологической активности и безопасности. Полученные результаты создают основу для дальнейшего продвижения данного класса соединений как перспективных кандидатов для разработки новых лекарственных препаратов	ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «РОССИЙСКИЙ НАЦИОНАЛЬНО-ИССЛЕДОВАТЕЛЬСКИЙ МЕДИЦИНСКИЙ УНИВЕРСИТЕТ ИМ. Н.И. ПИРОГОВА» МИНИСТЕРСТВА ЗДРАВООХРАНЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ	Поиск, дизайн и доклиническая оценка кандидатов

Наименование проекта	Описание проекта	Автор проекта	Этап разработки лекарств
Разработка соединений — потенциальных ингибиторов SARS-CoV-2 из панели S, N-содержащих соединений гетероцикличес-кого ряда (заключитель-ный)	В отчете представлены результаты исследования, направленного на синтез соединений — потенциальных ингибиторов SARS-CoV-2 из панели S, N-содержащих соединений гетероциклического ряда. Представлены результаты виртуального скрининга, молекулярного моделирования, синтеза и первичной биологической оценки соединений. Определены наиболее перспективные структуры для дальнейшего исследования в качестве потенциальных ингибиторов SARS-CoV-2	ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»	Поиск потенциаль-ных ингибиторов и первичный отбор кандидатов
Разработка вычислительных методов для идентификации вторичных метаболитов растений с использованием данных масс-спектрометрии (заключитель-ный)	Вторичные метаболиты растений лежат в основе многих лекарственных препаратов, включая антибиотики, иммунодепрессанты, витамины и другие. Современные технологии масс-спектрометрии позволяют за короткое время произвести сканирование образцов с тысячами метаболитов, получая в результате огромные объемы данных высокого разрешения. Несмотря на то что эти данные представляют собой перспективный источник для открытия новых соединений, их интерпретация остается узким местом, требующим развития соответствующих вычислительных методов. В рамках данного проекта были разработаны два новых подхода к высокопроизводительному анализу данных тандемной масс-спектрометрии вторичных метаболитов растений. Мы показали, что по точности работы наши методы соответствуют или превосходит все существующие аналоги и при этом обрабатывают данные на несколько порядков быстрее, что позволило применить их к анализу миллионов масс-спектров против базы с сотнями тысяч вторичных метаболитов. Кроме того, созданные программные инструменты обладают уникальной на данный момент функциональностью и позволяют находить в спектрах не только известные соединения, но и их новые варианты, не представленные в базе данных. Вычислительные методы были созданы в тесном сотрудничестве опытных разработчиков научного программного обеспечения и их конечных пользователей — исследователей вторичных метаболитов. Апробация	ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»	Поиск и идентифика-ция биологическ и активных соединений на ранней стадии

Наименование проекта	Описание проекта	Автор проекта	Этап разработки лекарств
	методов на больших объемах метаболомных данных показала их эффективность как для вторичных метаболитов растительного, так и микробного (бактерии, грибы) происхождения, таким образом расширяя охват использования созданных инструментов. Оба подхода доступны в виде приложения для командной строки и удобного веб-сервиса на GNPS, крупнейшей международной платформе по анализу масс-спектров вторичных метаболитов (Wang et al, Nature Biotechnology, 2016). Мы рассчитываем, что созданные методы будут активно использоваться другими научным группами, изучающими вторичный метаболизм, и помогут быстрее находить потенциально полезные биоактивные вещества природного происхождения		
Отчет о научно-исследовательской работе по теме «Анализ больших мультидисциплинарных данных для оптимизации создания и применения фармакотерапевтических средств в условиях коронавирусных эпидемий» (заключительный)	Целью проекта является разработка информационно-вычислительной системы (ИВС) на основе анализа развития ситуации с SARS-CoV-2/COVID-19, что позволит повысить эффективность репозиционирования известных лекарственных препаратов, создания новых фармакологических веществ и оптимизировать применение фармакотерапевтических средств с учетом индивидуальных особенностей и сопутствующих заболеваний пациентов. Нами реализована ИВС AntiCOVID-19, представленная в сети Интернет в виде свободно доступного портала AntiCOVID-19 (*http://www.way2drug.com/anticovid/*). Информационные компоненты ИВС: прошедшие экспертную оценку сведения о заболевании; данные о биологических процессах в организме человека, на которые влияет коронавирусная инфекция SARS-CoV-2; перечень молекулярных мишеней, воздействие на которые может блокировать развитие патологического процесса и/или снизить риск осложнений при COVID-19; перечень препаратов, рекомендуемых и исследуемых в качестве «кандидатов» для терапии COVID-19; ссылки на веб-ресурсы, связанные с данной предметной областью. Вычислительные компоненты ИВС: модуль построения гипотез для репозиционирования	ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ НАУЧНОЕ УЧРЕЖДЕНИЕ «НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ БИОМЕДИЦИНСКОЙ ХИМИИ ИМЕНИ В.Н. ОРЕХОВИЧА»	Репозиционирование и приоритизация кандидатов

Наименование проекта	Описание проекта	Автор проекта	Этап разработки лекарств
	лекарств на основе оценки структурного сходства по отношению к «препарату сравнения», модули прогнозирования профилей антикоронавирусной активности на основе классификационных моделей взаимосвязи «структура – активность» и модуль прогнозирования видов биологической активности, ассоциированных с фармакотерапевтическими эффектами, способствующими нормализации мультикомпонентных патологических процессов при COVID-19. Разработаны оригинальные методы интеллектуального анализа текстов для обработки слабо структурированной информации из научных публикаций по проблематике COVID-19 с целью актуализации информационного компонента ИВС AntiCOVID-19. Показана возможность извлечения терминов, соответствующих химическим соединениям с потенциальной активностью против белков коронавируса SARS-CoV-2, с применением оригинального подхода на основе мульти n-грамм, без использования промежуточных преобразований текстов обучающей, тестовой и целевой выборок в векторные представления. Применение разработанного метода позволяет осуществить распознавание новых наименований химических соединений, включая те, которые не встречались в выборках, используемых для обучения при построении модели. Реализована подсистема отбора фармакологических веществ, отвечающих заданному спектру биологической активности. С учетом новых опубликованных в литературе данных проведена актуализация обучающих выборок, на основе которых построены классификационные модели. Усовершенствован модуль генерации гипотез по репозиционированию лекарств для терапии COVID-19 на основе оценки структурного сходства с использованием MNA и QNA дескрипторов. Разработан подход к определению новых показаний к применению существующих лекарств на основе анализа их влияния на профили генной экспрессии в клеточных линиях человека, который позволяет		

Наименование проекта	Описание проекта	Автор проекта	Этап разработки лекарств
	осуществлять прогноз антикоронавирусной активности лекарственных препаратов, исходно введенных в медицинскую практику для терапии иных заболеваний. Дистанционным информационным центром по сложным вопросам фармакотерапии у пациентов с новой коронавирусной инфекцией COVID-19 «ФармаCOVID» (*https://rmapo.ru/pharmacovid.html*) проведена апробация ИВС AntiCOVID-19. Подготовлены рекомендации для врачей по применению различных препаратов (умифеновир, циклофосфамид, риамиловир и др.) для терапии COVID-19. Проведено усовершенствование ИВС AntiCOVID-19 с учетом замечаний клинических фармакологов. На основе накопленного за время пандемии COVID-19 опыта сформулированы предварительные рекомендации по ключевым направлениям развития фундаментальных и прикладных исследований и разработок, проведение которых позволит снизить негативные последствия новых биогенных угроз в будущем. Сопоставление полученных в ходе реализации проекта результатов с опубликованной в литературе информацией показывает, что данные разработки находятся на современном мировом уровне. Полученные результаты представлены в девяти статьях, включая аналитический обзор в Chemical Society Reviews (Q1, ИФ=60.615) и статью в Journal of Cheminformatics (Q1, ИФ=8.489). Результаты доложены и обсуждены на трех международных и шести российских (две с международным участием) конференциях и симпозиумах (шесть приглашенных пленарных и ключевых докладов)		
Отчет по заданию на выполнение государственных работ в сфере научной деятельности № FZSR-2023-0002 по теме «Полифенолы растений СФО:	Целью данной работы является анализ молекулярной и пространственной структуры полифенолов, выделенных из экстрактов ягод и растений Сибири, а также характеристика их биофункциональных свойств и токсикологических показателей безопасности на модельных системах in vivo. В ходе исследования проводились оценка структуры, анализ биологических эффектов и токсикологических	ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «КЕМЕРОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»	Ранняя доклиническая разработка биологически активных веществ

Наименование проекта	Описание проекта	Автор проекта	Этап разработки лекарств
оценка молекулярной и пространственной структуры веществ, характеристика биофункциональных свойств и токсикологических показателей безопасности на модельных системах in vivo» (заключительный)	характеристик на экспериментальных моделях. Полученные результаты могут быть использованы при дальнейшем исследовании природных соединений как основы для создания лекарственных и профилактических средств		
Репозиционирование лекарств для терапии COVID-19	Целью работы является обновление и пополнение информации в информационно-вычислительной платформе по репозиционированию лекарств для терапии COVID-19, а также подготовка аналитических материалов для органов государственной власти и участников российских проектов по поиску антикоронавирусных препаратов. В рамках работы проводился анализ литературных и экспериментальных данных, обобщение сведений о потенциальных препаратах-кандидатах, оценка их фармакологических свойств, а также развитие информационных ресурсов для поддержки принятия решений. Результаты проекта обеспечивают информационно-аналитическую поддержку участников российских проектов по поиску антикоронавирусных препаратов для терапии COVID-19	ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ НАУЧНОЕ УЧРЕЖДЕНИЕ «НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ БИОМЕДИЦИНСКОЙ ХИМИИ ИМЕНИ В.Н. ОРЕХОВИЧА»	Репозиционирование и отбор кандидатов для последующей проверки
Разработка методологии компьютерного поиска фармакологически активных соединений на основе множественного докинга и технологии искусственных нейронных сетей по теме ВАЛИДАЦИЯ РАЗРАБОТАН-	Объектом исследования являются: нейросетевая модель зависимости ГАМКА-антагонистической активности известных соединений от спектра их энергий множественного докинга; нейросетевая модель зависимости RAGE-ингибирующей активности известных соединений от спектра их энергий множественного докинга; данные по энергиям множественного докинга и градированному уровню RAGE-ингибирующей активности 183 известных соединений; экспериментальная 3D-модель α1-β2-γ2 подтипа ГАМКА-рецептора Homo sapiens, PDB-код 6D6T; 2D-структуры и субстанции пяти новых	ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «ВОЛГОГРАДСКИЙ ГОСУДАРСТВЕННЫЙ МЕДИЦИНСКИЙ УНИВЕРСИТЕТ» МИНИСТЕРСТВА ЗДРАВООХРАНЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ	Методология компьютерного поиска и первичный отбор кандидатов

Наименование проекта	Описание проекта	Автор проекта	Этап разработки лекарств
НОЙ МЕТОДОЛОГИИ НА ПРИМЕРЕ НЕСКОЛЬКИХ ВИДОВ ТАРГЕТНОЙ ФАРМАКОЛО-ГИЧЕСКОЙ АКТИВНОСТИ	производных хиназолин-2,4(1H,3H)-диона, шести новых производных 1,2,3,4-тетрагидропиразино(1,2-а)бензимидазола и референсного препарата диазепам; 78 белых беспородных мышей-самцов IC (CD-1) весом 21–24 гр. Цель работы — валидация разработанной методологии компьютерного поиска фармакологически активных соединений на основе множественного докинга и технологии искусственных нейронных сетей на примере нескольких видов таргетной фармакологической активности. В процессе работы проводили: прогноз с использованием нейросетевой модели на основе множественного докинга ГАМКА-антагонистической активности пяти новых производных хиназолин-2,4(1H,3H)-диона; экспериментальное тестирование на психотропную активность пяти новых производных хиназолин-2,4(1H,3H)-диона; прогноз с использованием нейросетевой модели на основе множественного докинга ГАМКА-антагонистической активности шести новых производных 1,2,3,4-тетрагидропиразино(1,2-а)бензимидазола; экспериментальное тестирование на психотропную активность шести новых производных 1,2,3,4-тетрагидропиразино(1,2-а)бензимидазола; обучение ранее найденных нейросетевых моделей зависимости RAGE-ингибирующей активности от спектра энергий множественного докинга химических соединений в режиме сэмплинга. Расчеты выполняли с применением параллельных вычислений с использованием суперкомпьютерной техники. Экспериментальное тестирование психотропной активности новых веществ выполняли с использованием установки «Открытое поле» с видеорегистрацией. В результате исследования впервые выполнен прогноз с использованием нейросетевой модели на основе множественного докинга ГАМКА-антагонистической активности и проведено экспериментальное тестирование на психотропную активность пяти новых производных хиназолин-2,4(1H,3H)-диона (точность прогноза составила 100%) и шести новых		

Наименование проекта	Описание проекта	Автор проекта	Этап разработки лекарств
	производных 1,2,3,4-тетрагидропиразино(1,2-а)бензимидазола (точность прогноза составила 83%), общая точность нейросетевого прогноза психотропной активности для 11 соединений составила 91%; выполнено обучение в режиме сэмплинга нейросетевых моделей зависимости RAGE-ингибирующей активности от спектра энергий множественного докинга химических соединений, существенно повышена точность прогноза; доказана валидность и адекватность новой разработанной методологии компьютерного поиска фармакологически активных соединений на основе множественного докинга и технологии искусственных нейронных сетей. Полученные данные в дальнейшем будут использованы для разработки методологии компьютерного поиска фармакологически активных соединений на основе множественного докинга и технологии искусственных нейронных сетей различной архитектуры, в том числе сверточных и нейросетей глубокого обучения		
Вычислительно доступный виртуальный скрининг лигандов для липидных GPCR-рецепторов (2 этап, заключительный)	Данный проект посвящен исследованиям в области рациональной разработки лекарств на основе структуры — современного подхода, который призван значительно уменьшить временные и денежные затраты на поиск новых лекарственных средств. Одним из этапов рациональной разработки лекарств является виртуальный скрининг ультрабольших (>100 млн соединений) виртуальных библиотек малых молекул. Идея использования структуры белка-мишени на этом этапе состоит в том, чтобы без оптимизации начальных низкомолекулярных соединений сразу найти лиганды, связывающиеся с целевой биомолекулой с хорошей аффинностью, что существенно облегчает дальнейшую проверку молекул. В последние три года такой подход показал свою пригодность при разработке лекарств к известным мишеням, таким как допаминовый рецептор D4 и мелатониновые рецепторы. Однако до сих пор набор мишеней, к которым реально может быть применен этот метод, сильно ограничен. На данный момент нет прямых	ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ (НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)»	Виртуальный скрининг и отбор кандидатов для экспериментальной валидации

Наименование проекта	Описание проекта	Автор проекта	Этап разработки лекарств
	свидетельств об использовании полученных за последние годы структур GPCR в рациональной разработке лекарственных препаратов к ним. Мы ставили перед собой амбициозную цель снять некоторые из ограничений, связанных с использованием методов молекулярного докинга и классических методов МО на размеченных данных при поиске потенциальных лигандов к целевым рецепторам. В первый год исследований, мы значительно продвинулись в построении алгоритма ускоренного докинга, а также в построении консенсусного, то есть основанного на умном усреднении нескольких независимых методов, метода построения скора для МО. Валидацию полученных методов проводили на опубликованной с нашим участием структуре липидного GPCR (S1P5), который вовлечен в патогенез целого ряда заболеваний — аутоимунных и нейродегенеративных (например, болезнь Альцгеймера и Хантингтона), гранулярную лейкемию, рассеянный склероз — и является перспективной фармакологической мишенью. Кроме того, появление нейросети AlphaFold, способа предсказания трехмерных структур белков, принципиально превосходящего по качеству доступные ранее методы, скорректировало работу: мы решили исследовать возможность использования сгенерированных AlphaFold структур для выполнения задач проекта. В результате этих исследований выяснили, что модели белков-мишеней, полученные при помощи AlphaFold, недостаточно точны для использования их в качестве шаблонов для вычислительного скрининга потенциально активных молекул. Тесты проводили с использованием структуры S1P5 рецептора, неизвестной на момент публикации AlphaFold. Результаты тестов показали, что структуры, сгенерированные нейросетью, не подходят для поиска селективных лигандов и не позволяют отделить активные молекулы от неактивных. Мы провели масштабный поиск параметров активного обучения в рамках построения системы по ускоренному скринингу. Оказалось, что для этой задачи лучше		

Наименование проекта	Описание проекта	Автор проекта	Этап разработки лекарств
	всего подходят простейшие алгоритмы МО (линейная регрессия с регуляризацией), обученные на небольших наборах данных. Кроме того, выяснилось, что алгоритмы, не обладающие памятью, то есть использующие каждый раз новые молекулы, показывают себя наравне с теми, которые каждый раз увеличивают размер тренировочного набора данных. Эти результаты позволяют построить простую и эффективную по вычислительным ресурсам систему виртуального скрининга молекул. Построение консенсусного алгоритма скоринга молекул, нацеленного на липидные GPCR, оказалось более трудоемким, чем предполагалось в начале проекта. Удалось построить консенсусные модели для двух самых больших наборов данных для рецепторов CysLT1 и A2A. В рамках отдельных рецепторов эти модели показали улучшение качества скрининга. Однако при тестировании полученных моделей на наличие обобщающей способности (выучивают ли они какую-то общую для всех липидных рецепторов информацию) выяснилось, что при обучении обобщения не происходит. Анализ имеющегося набора данных и алгоритма МО показали, что проблема, скорее всего, состояла в недостаточном объеме данных. В связи с этим план на второй год проекта был скорректирован с учетом необходимости составления собственного набора данных. По итогам первого года проекта часть полученных результатов была опубликована в журнале Nature Communications (п. 1 перечня публикаций). Кроме того, результаты были представлены на двух тематических конференциях. В течение второго года проекта работа была сконцентрирована на нескольких ключевых направлениях, каждое из которых имеет фундаментальное научное значение и потенциал в области вычислительной биологии и рациональной разработки лекарств. Наша работа в области ускоренного молекулярного докинга позволила продемонстрировать эффективность этого метода на общепринятых наборах данных. Для более широкого применения		

Наименование проекта	Описание проекта	Автор проекта	Этап разработки лекарств
	данного подхода мы разработали и внедрили инфраструктуру в алгоритм докинга Molsoft ICM. Алгоритм демонстрирует сходную работоспособность с методами, использующими глубокое обучение: он позволяет извлекать до 70% топ-0,5% всех лигандов (согласно оценке энергии связывания) после скрининга 2% лигандов, отобранных в результате скрининга, таким образом достигая ускорения в 50 раз и пропорционально уменьшая вычислительные ресурсы, требуемые для скрининга. Это возможно благодаря простому базовому алгоритму активного обучения — линейной регрессии на 2048 бинарных признаках с набором данных для обучения в 20 000 точек. Важно отметить, что обучение и применение данного алгоритма занимают менее 10 минут на наборе данных, включающем 100 млн лигандов, и имеют большой потенциал к расширению на ультрабольшие химические библиотеки. Вторым важным аспектом нашей работы стала ручная разметка данных базы CHEMBL. Мы смогли собрать значительные объемы данных для 21 липидного GPCR, включая более 600 вручную размеченных аннотаций экспериментов (по тому, позволяет ли эксперимент определить агонистическую активность лиганда, антагонистическую, или это нельзя заключить из описания), и свыше 20 000 лигандов с размеченной модальностью (агонист/антагонист). Однако, учитывая последние достижения в области ИИ, было принято решение отложить публикацию существующего набора данных в престижном научном журнале, расширив его за счет включения всех доступных экспериментальных данных для GPCR. Это планируется сделать с использованием методов ИИ, таких как ChatGPT, с валидацией на уже размеченных вручную данных. Предварительные результаты показали высокий потенциал данного подхода. Он позволит расширить лейблы модальности на миллионы лигандов, существенно увеличив набор данных по модальности лигандов для GPCR. Валидация отобранных лигандов проводилась при использовании структуры S1P5		

Наименование проекта	Описание проекта	Автор проекта	Этап разработки лекарств
	рецептора, полученной с нашим участием. Следует отметить, что S1P5 рецептор является важным объектом исследований в области фармакологии — он играет важную роль в регуляции различных биологических процессов. Применение структуры данного рецептора позволило более точно и целенаправленно отбирать потенциально активные лиганды, что является крайне важным при рациональной разработке лекарств. Мы разработали и успешно применили пайплайн для виртуального скрининга лигандов, основанный на структуре S1P5 рецептора, включающий передовые методы рационального виртуального скрининга. Благодаря этому, из огромных баз данных, содержащих более 6 млн соединений, было отобрано 100 потенциально значимых лигандов, имеющих высокую вероятность стать селективными лекарственными средствами, направленными на S1P5 рецептор. Успешность экспериментальной валидации таких лигандов повышается за счет того, что мы, учитывая особенности трехмерной структуры белка, более эффективно провели их отбор. Выбранные в результате виртуального скрининга лиганды будут проверены экспериментально с помощью клеточных тестов in vitro. Этот этап экспериментов играет ключевую роль в подтверждении потенциальной активности этих соединений и представляет значительный интерес для дальнейших исследований и разработки новых лекарственных препаратов. В целом, данная работа в отчетный период была ориентирована на развитие методов анализа данных и их применение в практических экспериментах, что расширило знания в области молекулярной биологии и фармакологии и может быть использовано в фармацевтической промышленности при разработке новых лекарственных препаратов. систему виртуального скрининга молекул		

Источник: составлено на основе данных Гиснаука за 2020–2025 гг.

Приложение 2

Проект — Концепция использования технологий ИИ и машинного обучения в жизненном цикле лекарственных средств

СОГЛАШЕНИЕ О ПРИНЦИПАХ ИСПОЛЬЗОВАНИЯ ТЕХНОЛОГИЙ МАТЕМАТИЧЕСКОГО МОДЕЛИРОВАНИЯ, МАШИННОГО ОБУЧЕНИЯ И ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЖИЗНЕННОМ ЦИКЛЕ ЛЕКАРСТВЕННЫХ СРЕДСТВ

1. Преамбула

Настоящее соглашение определяет единые принципы и подходы к разработке, внедрению и применению технологий математического моделирования (ММ), машинного обучения (МО) и искусственного интеллекта (ИИ) на всех этапах жизненного цикла лекарственных средств — от поиска новых молекул до вывода из оборота. Стороны соглашения руководствуются необходимостью обеспечения безопасности, эффективности и качества лекарственной терапии при одновременном стимулировании инноваций и гармонизации с международными регуляторными практиками.

2. Основополагающие принципы

2.1. Принцип технологической нейтральности

Регулирование основывается на оценке результатов и процессов, а не на предписании конкретных технологий или алгоритмов. Критерием приемлемости является научная обоснованность и валидность метода, а не его техническая реализация.

2.2. Принцип пропорциональности регуляторного подхода

Требования к валидации и верификации технологии ММ, МО и ИИ определяются в зависимости от рисков их применения и значимости принимаемых на их основе решений.

2.3. Принцип качества данных

Надежность системы напрямую зависит от качества и репрезентативности данных, использованных для ее обучения и валидации. Обеспечение качества данных является обязательным, но недостаточным условием. Окончательное подтверждение надежности системы достигается путем ее строгой экспериментальной валидации.

2.4. Принцип прозрачности, объяснимости и экспериментальной проверяемости

Процессы и решения, принимаемые с использованием технологий ММ, МО и ИИ, должны быть полностью документированы и научно обоснованы. Результаты и выводы, полученные с их применением, должны быть подвергнуты экспериментальной проверке в объеме, адекватном для подтверждения их надежности и точности в рамках заявленного контекста использования. Все методы и данные должны быть доступны для понимания и аудита со стороны регулирующих органов. Для систем «черного ящика» должны применяться методы объяснимого ИИ (XAI)

2.5. Принцип непрерывного мониторинга

Системы ИИ должны находиться под постоянным мониторингом на протяжении всего жизненного цикла с регулярной оценкой их эффективности и безопасности, а также процедуры его обновления и переобучения в случае их деградации («дрейфа модели») или появления новых данных.

2.6. Принцип управления рисками и валидации

Любое применение технологий ММ, МО и ИИ в жизненном цикле лекарственного средства должно сопровождаться оценкой рисков, пропорциональной значимости принимаемых решений. Любые используемые модели должны проходить строгую валидацию, включая оценку на независимых наборах данных, анализ устойчивости и исследование влияния смещений на результаты.

3. Применение ИИ на этапах жизненного цикла ЛС

3.1. Открытие и разработка прототипа ЛС

- Использование ИИ для идентификации мишеней, дизайна молекул, прогнозирования физико-химических и фармакологических свойств препаратов и установления их механизма действия.
- Требование: строгое следование установленным стандартам и рекомендациям по применению технологий ММ, МО и ИИ; экспериментальное подтверждение in vitro/in vivo ключевых предсказаний.

3.2. Доклинические исследования

- Анализ экспериментальных данных (омиксные данные, гистологические изображения, in vivo токсикология/фармакология и др.); трансляционные исследования и дизайн стратегии клинической разработки.
- Требование: строгое следование установленным стандартам и рекомендациям по применению технологий ММ, МО и ИИ; ортогональное экспериментальное подтверждение (in vitro/in vivo) ключевых выводов, сделанных на основе применения технологий ММ, МО и ИИ, валидация на независимых наборах данных.

3.3. Клинические разработки

- Стратификация пациентов; разработка суррогатных биомаркеров и цифровых конечных точек КИ; рациональный дизайн клинических исследований (оптимизация дизайна, платформенные КИ, КИ с историческим и виртуальным контролем и др.); количественная фармакология; дополнительное доказательство эффективности и безопасности КИ; мониторинг безопасности.
- Требование: строгое следование установленным стандартам и рекомендациям по применению технологий ММ, МО и ИИ; соответствие GCP; использование единых стандартов обработки и хранения данных КИ для обеспечения их репрезентативности;

3.4. Регистрация и экспертиза

- Использование технологий ММ, МО и ИИ для анализа данных регистрационного досье; регистрация медицинских изделий, включающих элементы на основе технологий ММ, МО и ИИ.
- Требование: предоставление исчерпывающей технической и эксплуатационной документации, включая научное обоснование

применяемых методов, а также доказательства воспроизводимости и валидации полученных результатов в рамках заявленного контекста использования.

3.5. Пострегистрационный этап

- Фармаконадзор на основе ИИ, анализ данных реальной клинической практики
- Требование: использование единых стандартов обработки и хранения данных КИ для обеспечения их репрезентативности; интеграция с существующими системами безопасности и наличие утвержденного плана управления рисками (Risk Management Plan, RMP), включающего мониторинг работы ИИ-моделей.

3.6. Вывод из оборота

- Использование технологии ММ, МО и ИИ для анализа данных об отзыве препаратов
- Требование: обеспечение преемственности данных и решений

4. Механизмы реализации

4.1. Научные консультации

Создание системы научного консультирования на ранних этапах разработки для обсуждения планов применения ИИ, включая выбор методов валидации и управления рисками.

4.2. Регуляторные песочницы

Организация специальных режимов для тестирования инновационных ИИ-решений в контролируемых условиях, с возможностью адаптации регуляторных требований.

4.3. Совместные рабочие группы

Формирование рабочих групп с участием всех заинтересованных сторон для разработки детальных рекомендаций и стандартов.

4.4. Гармонизация требований

Активное участие в международных инициативах по гармонизации регуляторных подходов к ИИ в фармацевтике. Разработка и применение методов ИИ должны учитывать лучшие международные практики и перспективные стандарты (ISO/IEC, IEEE, NIST).

4.5. Развитие регуляторных компетенций

Стороны обязуются способствовать развитию компетенций в области ИИ среди сотрудников регуляторных органов, включая обучение, обмен опытом с международными регуляторами и привлечение внешних экспертов.

5. Механизм разработки и принятия самостоятельных документов

5.1. На основании настоящего Рамочного соглашения Стороны инициируют разработку системы детализирующих отраслевых руководств, стандартов и регламентов.

5.2. Координацию работ по разработке осуществляет Совет по внедрению ИИ в фармацевтике (или Рабочая группа), сформированный из представителей всех Сторон.

5.3. Для разработки каждого отдельного документа по конкретному направлению Совет вправе формировать тематические экспертные группы с привлечением ведущих российских и международных специалистов.

5.4. Все разрабатываемые документы должны проходить этап публичного обсуждения и пилотного тестирования с отраслевым сообществом перед их окончательным принятием.

5.5. Разработанные и утвержденные документы являются логическим развитием настоящего Соглашения и обязательны к применению всеми Сторонами в пределах их компетенции.

6. Заключительные положения

Стороны обязуются способствовать созданию и обеспечить доступ к анонимизированным регуляторным наборам данных (датасетам) для целей обучения и валидации алгоритмов, в соответствии с законодательством о персональных данных и интеллектуальной собственности. В целях развития доверия к применяемым методам ИИ, приветствуется участие в международных инициативах по бенчмаркингу и сравнительной оценке алгоритмов.

Настоящее соглашение является основой для разработки детальных отраслевых рекомендаций и руководств. Стороны обязуются ежегодно пересматривать и актуализировать положения соглашения с учетом развития технологий и международного опыта.

Соглашение открыто для присоединения других заинтересованных организаций и ведомств.

Приложение к Соглашению о принципах использования технологий искусственного интеллекта в жизненном цикле лекарственных средств

ГЛОССАРИЙ ОСНОВНЫХ ТЕРМИНОВ

Настоящий Глоссарий определяет основные термины, используемые в Соглашении и разрабатываемых на его основе документах, в целях обеспечения единообразного толкования.

Алгоритм (Algorithm) — конечная последовательность шагов или инструкций, предназначенных для решения определенной задачи или класса задач.

Алгоритмическая предвзятость (Algorithmic Bias) — систематическая и воспроизводимая ошибка в работе системы искусственного интеллекта, приводящая к несправедливым или дискриминационным результатам, часто возникающая из-за недостаточной репрезентативности или смещений в обучающих данных.

Большие данные (Big Data) — объемные, сложные, быстро обновляемые и многообразные по своей природе наборы данных, для обработки и анализа которых требуются специальные технологические подходы и инструменты, включая методы ИИ.

Валидация (Validation) — подтверждение на основе объективных свидетельств того, что требования, предъявляемые к конкретному использованию модели ИИ, выполнены. В контексте ИИ — оценка того, насколько модель пригодна для достижения поставленной цели в реальных условиях.

Верификация (Verification) — подтверждение на основе объективных свидетельств того, что установленные требования к модели ИИ выполнены. В контексте ИИ — проверка корректности технической реализации модели (например, «правильно ли мы построили модель?»).

Данные для обучения (Training Data) — набор данных, используемый для обучения (настройки параметров) модели машинного обучения.

Данные для тестирования (Test Data) — независимый набор данных, не использовавшийся при обучении модели, применяемый для объективной оценки ее итоговой производительности и способности к обобщению.

Данные для валидации (Validation Data) — набор данных, используемый для настройки гиперпараметров модели и промежуточной оценки в процессе обучения с целью предотвращения переобучения.

Дрейф модели (Model Drift) / Концептуальный дрейф (Concept Drift) — изменение статистических свойств целевой переменной или взаимосвязей в данных с течением времени, приводящее к постепенному снижению прогнозной способности модели после ее развертывания.

Естественный язык (Natural Language) — язык, который развивался естественным путем, как средство общения между людьми (например, русский, английский).

Жизненный цикл модели ИИ (AI Model Lifecycle) — все этапы существования модели ИИ, включая проектирование, разработку, обучение, валидацию, внедрение, мониторинг, обновление и вывод из эксплуатации.

Задача (Task) — целенаправленная деятельность, для выполнения которой предназначена система ИИ.

Знание (Knowledge) — абстрактное или конкретное понимание субъекта или объекта.

Интеллект (Intelligence) — способность системы ИИ решать задачи.

Интерпретируемость (Interpretability) — свойство модели, позволяющее человеку понять причинно-следственные связи между входными данными и результатами работы модели.

Искусственный интеллект, ИИ (Artificial Intelligence, AI) — технология, основанная на компьютерных алгоритмах, предназначенная для решения когнитивных задач, традиционно ассоциирующихся с человеческим интеллектом, таких как обучение, анализ данных, восприятие окружающей среды и принятие решений. В контексте настоящего Соглашения под ИИ понимаются системы, способные к выполнению сложных задач в изменяющейся среде.

Качество данных (Data Quality) — совокупность характеристик данных, определяющих их пригодность для достижения поставленных целей, включая полноту, точность, актуальность, непротиворечивость и релевантность.

Контекст использования (Context of Use) — полное и четкое описание того, как и для какой конкретной цели предполагается использовать модель ИИ в процессе разработки, регистрации или мониторинга лекарственного средства.

Машинное обучение, МО (Machine Learning, ML) — подраздел искусственного интеллекта, представляющий собой класс методов, позволяющих компьютерным системам «обучаться» на основе данных (выявлять в

них закономерности) для решения задач без явного программирования на каждую из них.

Модель (Model) — упрощенное представление реальности, созданное для достижения определенной цели.

Математическое моделирование, ММ (Modeling & Simulation, M&S) — изучения реальных объектов, процессов или явлений путем создания их идеализированных описаний на языке математики, называемых математическими моделями.

Модель «черный ящик» (Black Box Model) — модель ИИ/МО, внутренняя логика и механизм принятия решений которой являются сложными для интерпретации человеком.

Нейронная сеть (Neural Network) / Искусственная нейронная сеть (Artificial Neural Network) — вычислительная модель, архитектура которой вдохновлена биологическими нейронными сетями мозга и состоит из взаимосвязанных узлов (нейронов), организованных в слои. Используется для решения широкого круга задач, таких как распознавание образов и прогнозирование.

Объяснимый искусственный интеллект, ОИИ (Explainable AI, XAI) — набор методов и инструментов, позволяющих понять и объяснить выводы и решения, принимаемые моделями ИИ, в том числе моделями типа «черный ящик». Критически важен для обеспечения доверия, проверки и принятия решений в регулируемых областях.

Обработка естественного языка, ОЕЯ (Natural Language Processing, NLP) — область искусственного интеллекта, занимающаяся взаимодействием компьютеров и человеческого языка, в частности, разработкой алгоритмов для анализа, понимания и генерации естественного языка.

Обучение с учителем (Supervised Learning) — вид машинного обучения, при котором модель обучается на размеченных данных, то есть на примерах, для которых известны правильные ответы (целевые переменные).

Обучение без учителя (Unsupervised Learning) — вид машинного обучения, при котором модель ищет скрытые закономерности или внутренние структуры в неразмеченных данных.

Обучающая выборка (Training Sample) — часть набора данных, используемая для обучения модели.

Переобучение (Overfitting) — ситуация, когда модель машинного обучения слишком точно подстраивается под обучающие данные, включая их случайный шум, и вследствие этого плохо работает на новых, ранее не виденных данных.

Предвзятость (Bias) — См. Алгоритмическая предвзятость.

Предсказание (Prediction) — результат, полученный при помощи модели.

Прогноз (Forecast) — предсказание будущих событий.

Репрезентативность данных (Data Representativeness) — свойство набора данных, означающее, что он адекватно и непредвзято отражает характеристики целевой популяции или явления, для изучения которых он предназначен.

Цифровой двойник (Digital Twin) — виртуальная динамическая модель физического объекта или процесса, которая постоянно обновляется данными с его датчиков и может использоваться для анализа, прогнозирования и оптимизации.

Человек в контуре (Human-in-the-Loop, HITL) — подход к проектированию систем ИИ, предполагающий активное участие человека в ключевых этапах

работы алгоритма, таких как подготовка данных, валидация результатов или принятие окончательных решений, особенно в ситуациях с высокими рисками.

Термины и определения в данном Глоссарии могут уточняться и дополняться по мере развития технологий и нормативно-правовой базы. При разработке детализированных руководств рекомендуются к использованию международные стандарты в области ИИ (ISO/IEC, IEEE, NIST).

9 785605 618140