



Бердунин В.А.  
Романова Т.Е.

## МЕДИЦИНСКИЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

РЕАЛЬНОСТЬ,  
ПЕРСПЕКТИВЫ,  
ИЛЛЮЗИИ  
И УГРОЗЫ

Монография

Федеральное медико-биологическое агентство  
Федеральное государственное бюджетное учреждение  
«Государственный научный центр Российской Федерации –  
Федеральный медицинский биофизический центр  
имени А.И.Бурназяна»  
Медико-биологический университет  
инноваций и непрерывного образования

# **МЕДИЦИНСКИЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ**

## **РЕАЛЬНОСТЬ, ПЕРСПЕКТИВЫ, ИЛЛЮЗИИ И УГРОЗЫ**

### **МОНОГРАФИЯ**

**Москва  
2025**

**УДК 004.8**

**ББК 5с51+16.6**

**Б48**

**Бердутин В.А., Романова Т.Е.** Медицинский искусственный интеллект. Реальность, перспективы, иллюзии и угрозы. Монография. М.: ФГБУ ГНЦ ФМБЦ им. А.И. Бурназяна ФМБА России, 2025. 186 с.

**Авторы:**

**Бердутин В.А.** – ФГБУ «Государственный научный центр Российской Федерации – Федеральный медицинский биофизический центр имени А. И. Бурназяна» ФМБА России

**Романова Т.Е.** – ФГБОУ ВО «Приволжский медицинский исследовательский университет» Минздрава России, Нижний Новгород

**Рецензенты:**

**Абаева Ольга Петровна** – доктор медицинских наук, доцент, профессор кафедры социологии медицины, экономики здравоохранения и медицинского страхования Первого МГМУ им. И. М. Сеченова

**Кром И.Л.** – доктор медицинских наук, профессор кафедры общественного здоровья и здравоохранения с курсами (с курсами правоведения и истории медицины) Саратовского государственного медицинского университета им. В.И. Разумовского

В глобальном масштабе прямо на наших глазах происходит коренная трансформация сферы здравоохранения. Последние несколько лет стали переломными, если рассуждать о количестве новых направлений, появившихся лечебно-диагностических методик и внедрении цифровых платформ. Цифровая медицина использует информационно-коммуникационные технологии, чтобы уже сегодня успешно решать многочисленные проблемы, связанные с обеспечением качества и доступности медицинской помощи. Стремительное развитие нейросетей и искусственного интеллекта (ИИ) предоставляет медицинским работникам широкие возможности при прогнозировании течения заболеваний и расчете рисков для здоровья пациентов. Производители медицинских гаджетов предлагают потребителям широкий ассортимент программного обеспечения и товаров с элементами ИИ. Носимые устройства подобного рода постепенно становятся обыденными средствами поддержки лечебного процесса и мониторинга состояния человека. Несмотря на грандиозные успехи в деле применения ИИ в медицине, врачебное сообщество весьма обеспокоено некоторыми сложно разрешаемыми проблемами, связанными со слишком быстрым и повсеместным использованием указанных цифровых платформ. Глубокая нейронная сеть представляет собой крайне сложно устроенную компьютерную программу, состоящую из большого количества внутренних скрытых слоев с настраиваемыми параметрами. Чем сложнее устроена нейросеть и чем больше вычислительных операций она выполняет, тем труднее разобраться, что происходит в ее внутренних слоях. Функционирование систем ИИ в формате чёрного ящика делает объяснение результатов их работы весьма нетривиальной задачей. В будущем наверняка потребуются дополнительные исследования в области оценки надежности и интерпретации процессов принятия решений в этих системах, что, в

первую очередь, коснётся нейросетей самых последних поколений. Поэтому наша монография нацелена на то, чтобы дать заинтересованной аудитории развёрнутое представление о реальной ситуации, перспективах и рисках применения медицинского искусственного интеллекта, избавив работников здравоохранения от досадных заблуждений и избытка иллюзий, которые и ныне могут встречаться в профессиональной среде. Помимо обобщения соображений, выводов и рекомендаций, содержащихся в монографии, там приводятся конкретные примеры из различных областей медицины: радиологии, кардиологии, хирургии, офтальмологии, патоморфологии и т. д., которые убедительно иллюстрируют достижения и потенциал будущих направлений развития медицинского ИИ. Содержание книги должно быть интересным не только врачам и медработникам среднего звена, но и разработчикам медицинского программного обеспечения, служащим государственных органов управления здравоохранением и бюро медико-социальной экспертизы, широкому кругу представителей экспертного сообщества и страховых медицинских компаний, лицам, занятым клиническими исследованиями, социологами, фармацевтам, а также преподавателям и студентам медицинских университетов и колледжей.

## СОДЕРЖАНИЕ

Введение .....	6
1. Приложения искусственного интеллекта	
в здравоохранении .....	15
• Клинические нейросетевые платформы .....	15
• Применение нейросетей в биомедицинских	
исследованиях .....	22
• Применение ИИ в сфере общественного	
здравоохранения .....	24
• Риски применения ИИ в здравоохранении .....	26
• Методологические аспекты оценки рисков	
медицинского искусственного интеллекта .....	40
2. Разработка, обучение и развёртывание платформ	
медицинского искусственного интеллекта .....	48
• Разработка платформ медицинского ИИ .....	48
• Обучение моделей ИИ .....	55
• Развертывание платформ медицинского ИИ	
в клинике .....	68
• Проблемные аспекты развёртывания платформ	
ИИ .....	74
• Организационное управление и эксплуатация	
платформ ИИ .....	78
• Клиническая валидация и мониторинг инструментов	
медицинского ИИ .....	85
• Интерпретируемость рекомендаций ИИ .....	91
3. Искусственный интеллект в здравоохранении:	
надежда, а не шумиха, обещание, а не опасность .....	95
• Зимы ИИ и парадоксы в моделях машинного	
обучения .....	95
• Проблемы и ключевые приоритеты в области	
стандартизации ИИ .....	102
• Конкуренция нейросетей в области	
дифференциальной диагностики опухолей мозга .....	108
• Ближайшие перспективы медицинского ИИ .....	116
4. Эмоциональный искусственный интеллект .....	125
• Эмоциональный интеллект в деятельности врача .....	125
• Эмоциональные вычисления .....	128
• Морально-правовой базис ЭИИ .....	135

• Нанотехнологии в области ИИ .....	138
5. Искусственный интеллект для здравоохранения от McKinsey .....	140
• Институт здравоохранения McKinsey .....	140
• Гибридный искусственный интеллект: сила цифровых технологий + сила людей .....	145
6. Нечеткие нейронные сети в медицине .....	148
• Родоначальник современной неклассической логики Н.А. Васильев .....	148
• Слияние нечеткой логики и нейронных сетей – апофеоз медицинского ИИ .....	151
• Проблематика эффекта нелинейности нейронных сетей в представлении выпускников цифровых кафедр медицинских вузов .....	162
Заключение .....	169
Определения и термины .....	171
Список литературных источников .....	175



## ВВЕДЕНИЕ

Отечественное здравоохранение с невероятной быстротой видоизменяется под давлением инновационных цифровых платформ, приобретших популярность под названиями «искусственный интеллект» (ИИ) (англ. «artificial intelligence», AI) и «машинное обучение» (англ. «machine learning», ML). Очевидно, что эти привлекающие нас фантастическими возможностями платформы имеют не только полезные качества, но и несут в себе серьёзные угрозы и риски. Несмотря ни на что, сегодня мы не в силах отказаться от тех преимуществ, которые они уже предоставляют в деле управления здоровьем населения с помощью глубокого обучения многоуровневых нейронных сетей, а также множества открытий, приведших к возросшему уровню точности диагностики и успехам в хирургической робототехнике. Оперирующие собираемыми отовсюду массивами данных, системы ИИ способны гарантированно обеспечить принятие оптимальных решений ключевым игрокам сферы здравоохранения, начиная от фармацевтической индустрии и заканчивая самими мелкими поставщиками медицинских услуг [3, 4].

В совместном докладе компаний IDC и Seagate (США) «Эпоха данных-2025» предполагается, что к середине 2-го десятилетия XXI в. общемировой объём данных достигнет 163 зеттабайт<sup>18</sup> D<sub>+</sub> то есть  $1,63 \cdot 10^{23}$  байт. К этому времени около 96% информационных транзакций между подключёнными к сети устройствами будет осуществляться в режиме онлайн, что фактически можно считать не чем иным, как *интернетом вещей* (англ. «internet of things», IoT). Создание гигантского массива данных, накапливаемых с триллионов единиц оборудования, прорывные решения в области облачных технологий и программных платформ – всё это предтечи неуклонно надвигающейся эры IoT и ИИ. В этом не приходится сомневаться, поскольку информация уже сегодня стала основным движущим фактором управления в большинстве отраслей российской экономики целом и в отрасли здравоохранения, в частности [2, 46].

В настоящее время мировая индустрия здравоохранения представляет собой крайне запутанную систему стоимостью более 2 трлн \$. Медицинская сфера, буквально раздавленная под гнётом колоссальных непроизводительных затрат и бюрократических ограничений, лихорадочно нащупывает пути выхода из кризиса, усугубившегося ещё в период пандемии COVID-19. В качестве выхода из сложившейся ситуации учёные рассматривают внедрение инноваций в области медицинских цифровых платформ, которые должны помочь в решении ключевых проблем повышения качества и эффективности системы здравоохранения. Согласно результатам международных исследований, применение платформ ИИ способно увеличить валовую прибыль фирм и организаций, связанных с медицинской деятельностью. В 2016 году стоимость европейского рынка ИИ составила 270 млн евро; прогнозируемый ежегодный прирост оценивали в 36%. Бюро промышленности и безопасности

США считает, что к 2025-2026 году рынок медицинского ИИ составит порядка 28 млрд \$, а доля рынка медицинской диагностики и визуализации займет в нём около 2,5 млрд \$ [25, 30, 58, 63, 137].

Основной мотивацией для проникновения ИИ в здравоохранение служат: перманентный рост затрат, острая необходимость их сокращения, проблема некачественной диагностики (до 30% исследований на поверку оказываются недостоверными либо неверно интерпретируемыми), а также стремление к стандартизации и автоматизации рутинных функций вплоть до создания самоуправляемых диагностических моделей. Дополнительным драйвером популярности ИИ стал нарастающий как снежный ком гигантский массив данных, поступающих от различных медицинских устройств и информационных систем. Безусловно, ИИ привлекает возможностью создавать дополнительную стоимость за счёт больших данных (англ. «big data», BD), поэтому, как следует из известного девиза клиники Мэйо (США) «doing more with less», призывающего снижать затраты, сохраняя эффективность, сегодня наступил как раз тот благоприятный момент, чтобы выгодно инвестировать лежащий мертвым грузом капитал в быстрорастущие медицинские стартапы. Американская международная консалтинговая компания со штаб-квартирой в Нью-Йорке McKinsey&Company заявила, что в настоящий момент в здравоохранении возможно автоматизировать до 36% функций, прежде всего на уровнях сбора и анализа данных. В результате на зарождающемся рынке медицинского ИИ уже работает целый ряд бизнес-моделей: продажа узкоспециализированных сервисов медицинским учреждениям, продажа дополнительных лицензий на сервисы через маркетплейсы, продажа стартапами инновационного программного обеспечения (ПО) крупным компаниям-интеграторам, занимающимся медицинскими информационными системами, иные всевозможные виды партнёрства между стартапами и известными промышленными лидерами. Все игроки рынка интеллектуального ПО прекрасно понимают, что «золотая жила» лежит прямо у них ногами, и тот, кто найдет её первым, окажется фаворитом [143, 145, 153].

Итак, медицинский ИИ начал своё уверенное движение в сторону революционных преобразований сферы здравоохранения. Ожидается, что совсем скоро он существенно улучшит доступность медицинской помощи, способы лечения пациентов, оптимизируя систему распределения ресурсов, тем самым помогая организациям здравоохранения функционировать более эффективно. Потенциал искусственного интеллекта, без сомнения, способен в корне изменить отрасль здравоохранения, помочь повысить точность и сократить время диагностики сложных для верификации заболеваний, обеспечить по-настоящему персонализированный подход к пациенту. Некоторые из основных применений медицинского ИИ включают количественную оценку медицинских изображений, автоматический анализ генетических данных, прогнозирование заболеваний, медицинскую робототехнику, телемедицину, создание



виртуальных врачей-консультантов и проч. Как ни странно, но мировой экономический кризис, вызванный пандемией коронавируса, послужил своеобразным катализатором разработки и внедрение множества клинических приложений ИИ [23, 24, 39, 48, 53].

В недавних исследованиях использовали доступность больших наборов медицинских текстовых данных для задач обработки естественного языка, связанных со здравоохранением, с использованием технических достижений, таких как преобразователи и контекстные вложения слов (2 технологии, которые помогают моделям учитывать окружающий контекст при интерпретации каждой части текста). В одном из исследований была представлена BioBERT модель, обученная на большом корпусе медицинских текстов, которая превзошла предыдущие достижения в выполнении задач на естественном языке, таких как ответы на биомедицинские вопросы. Подобные модели использовали для повышения производительности в задачах изучения биомедицинской литературы о взаимодействии лекарств друг с другом или об автоматизации рентгенологических отчетов. Также значительные объемы текстовых данных извлекали из социальных сетей и применяли для отслеживания крупномасштабных тенденций в области психического здоровья. В итоге достижения в обработке естественного языка открыли множество новых наборов данных и возможностей ИИ, хотя всё ещё остаётся немало ограничений ввиду сложности извлечения информации из длинных текстовых последовательностей [55, 57, 70].

Предоставление медицинской помощи во многих странах регулируется институтом страхования, который вот уже несколько десятилетий активно и успешно использует компьютерные технологии безопасности класса insurtech (англ. аббревиатура "insurance technology"). Здесь алгоритмы ИИ применяют для обработки претензий и администрирования платежей, поскольку страховщики и поставщики медицинских услуг должны проверять достоверность данных в настоящей лавине постоянно осуществляемых транзакций. Своевременное обнаружение проблем с кодированием и исправление ошибок экономят финансовые и временные ресурсы всех стейкхолдеров (англ. «stakeholder» заинтересованная сторона) [75, 80].

Компания KenSci из США пытается совместить BD и ИИ в целях прогнозирования клинических, финансовых и операционных рисков, используя существующую в доступных источниках информацию, чтобы предсказать различные сценарии развития ситуации на рынке медицинских услуг, например, сколько и каких людей может заболеть в ближайшее время, а также насколько это увеличит расходы медицинских организаций. KenSci сотрудничает со многими крупнейшими американскими корпорациями, включая Дженерал Электрик, KPMG, Allscripts и Microsoft. Цифровая платформа H2O.ai из Маунтин-Вью, Калифорния, способна анализировать данные, собираемые по всей системе здравоохранения США с тем, чтобы прогнозировать и оптимизировать её

деятельность. Она позволяет составлять точный прогноз загрузки отделений интенсивной терапии больниц, устанавливать уровни риска развития внутрибольничной инфекции у пациентов и в целом добиваться улучшения качества медицинской помощи. Используя подобные платформы, врачи могут точнее диагностировать сепсис, что, безусловно, позволяет значительно снижать уровень госпитальной летальности [89, 93, 94, 97].

Китайская компания из Шэньчжэня ICarbonX использует ИИ и ВД для тщательного изучения характеристик жизни пациентов, составляя цифровую модель их жизни. Компания надеется, что её ВД станут настолько мощными, что когда-нибудь смогут управлять всеми аспектами здоровья. ICarbonX считает, что её технологии смогут собрать достаточно данных для исчерпывающей классификации симптомов нозологий и разработки всех возможных вариантов их лечения [85].

Однако, мы не должны сбрасывать со счетов, что, как и в случае с любыми технологическими новшествами, медицинский ИИ имеет не только очевидные преимущества, но и несет с собой весьма специфические риски. Поэтому он нуждается в жестких ограничивающих нормативных рамках, которые бы учитывали все экономические и социально-этические последствия его использования. Обеспокоенность у профессиональной общественности и пациентов вызывают множественные проблемы, касающиеся вопросов обеспечения клинической безопасности, регулирования доступа, конфиденциальности, ответственности и др. Острой проблемой применения ИИ в медицине и биологии является техническая и юридическая сложность установления подотчетности лиц, так или иначе причастных к обработке данных и/или обслуживанию компьютерных систем [1, 5, 31, 51].

Нарастание популярности ИИ вызывает обоснованные опасения в отношении юридической и профессиональной ответственности, поскольку до сих пор неясно, должны ли разработчики, регулирующие органы, продавцы или поставщики медицинских услуг нести эту ответственность, если платформа допускает ошибки даже после тщательной клинической проверки. В настоящее время врачи подвергаются санкциям в случае отклонения от стандарта оказания медицинской помощи и причинения вреда здоровью пациента. Если врач скептически относится к ИИ, то он может проигнорировать его рекомендации, противоречащие стандартной практике, даже если эти рекомендации могут оказаться персонализированными и полезными для конкретного пациента. Однако если стандарты медицинской помощи изменятся так, что врачи будут регулярно использовать инструменты ИИ, у них появится мощный судебно-медицинский стимул следовать его рекомендациям. Конфиденциальные данные, как правило, должны собираться и использоваться с согласия пациента и, дабы не афишировать личные данные, целесообразно использовать стратегии анонимизации и агрегирования. В принципе личные данные пациентов

должны использоваться исключительно для их благополучия. Необходимо, чтобы все учреждения, работающие с данными пациентов, относились к ним ответственно, используя соответствующие протоколы безопасности. Между тем внедрение мер безопасности зачастую требует значительных ресурсов и усилий, а учреждения, которые не могут позволить себе достаточных инвестиций в этой области, рано или поздно оказываются в затруднительном положении, даже если их деятельность является общественно полезной. К тому же регулярная практика запросов на использование данных в разных учреждениях и проектах может затруднить получение добровольного информированного согласия, поскольку люди будут с недоверием относиться к повторяющимся требованиям о предоставлении персональных данных. Также остаются вопросы о регулировании ИИ в медицине и о том, как он может сокращать и создавать рабочие места во всей системе здравоохранения, затрагивая интересы в равной степени работодателей, врачей и пациентов [28, 29, 36, 60, 74].

Конкретные проблемы регуляторного характера будут неизбежны из-за необходимости обучения ИИ, поскольку цифровая платформа с течением времени накапливает всё новые и новые, возможно, «мусорные» данные, приспосабливаясь к популяционным изменениям социума. Тут также кроется вероятность возникновения ошибок. Традиционно регуляторы систем ИИ утверждают только один фиксированный набор параметров, но этот подход не учитывает необходимость обновления информации, поскольку входные данные и параметры постоянно меняются. По этой причине надзорные органы должны опираться на возобновляемые процедуры сертификации нейросетей. Не так давно была высказана идея о том, что для адаптивных систем ИИ, должна утверждаться не только исходная платформа, но и все процедуры её обновления с течением времени [78, 84].

Хотя ИИ расширяет возможности врача в принятии медицинских решений, одновременно он создаёт предпосылки к ограничению личной свободы и возникновению нежелательных обязательств. Поскольку платформы ИИ берут на себя больше и больше функций, проблема, с которой сталкивается здравоохранение, состоит в том, что клиницисты могут чересчур полагаться на нейросеть, констатируя при этом постепенное снижение собственных навыков и утрату доверительного контакта с больным. В свою очередь, разработчики платформ с ИИ приобретают избыточное влияние, которое обязаны использовать для создания безопасных и полезных людям систем. Поскольку принятие медицинских решений сильно зависит от во многом необъяснимых суждений ИИ, отдельные больные, утратив контроль над своим собственным состоянием, могут потерять всякое понимание происходящего. И в то же самое время пациент вдруг обременяется некими ранее не присущими ему обязанностями. Например, если интеллектуальные устройства навязчиво предлагают пациентам различные советы, то есте-

ственно ожидать, что те будут следовать этим рекомендациям, а, следовательно, возьмут на себя ответственность за возможные негативные последствия для собственного здоровья [83, 100, 110].

Еще один острый вопрос заключается в том, способен ли ИИ обеспечить равную доступность оказания медицинской помощи всем без исключения слоям населения, или же он только будет способствовать увеличению общественно-экономического неравенства и несправедливости по отношению к малообеспеченным социальным группам. Дело в том, что платформы ИИ имеют неприятное свойство масштабировать общественные настроения, замешанные на предрассудках и предубеждениях, которые в неявном виде всегда присутствуют в накапливаемых данных. Алгоритмы ИИ не могут оперировать недоступными им данными, поэтому обучение ИИ на наборах данных, не представляющих социально незащищенные и маргинализированные группы населения, приводит к системным ошибкам платформ, которые будут игнорировать интересы этих групп. В то же время цифровые платформы, которые преднамеренно нацелены на массовый учёт потребностей прекариата, иных экономически несостоятельных слоёв, мигрантов, национальных меньшинств и прочих ущемлённых в правах категорий населения, прямо рискуют закрепить в своих алгоритмах состояние неравенства, потому что осуществить этически справедливую дифференциацию столь разнотипной массы людей компьютерной программе, не обладающей критическим мышлением, в принципе невозможно [87, 132, 138].

Неправильная интерпретация не может не возникать из-за неудачных вариантов дизайна платформ, например, ошибочного выбора целевой метки. Так, в США было установлено, что алгоритм оценки риска, используемый для руководства принятием клинических решений для 200 млн пациентов, делал прогнозы с расовой предвзятостью, потому что белые пациенты, которым присваивался определённый балл риска, как правило, оказывались здоровее, чем афроамериканцы с таким же баллом. Это предубеждение оказалось в значительной степени связано с целевыми метками, используемыми в ML. Поскольку предназначенный для прогноза расходов на здравоохранение алгоритм ИИ был обучен тому, что чернокожие пациенты на протяжении исторически длительного периода получали самое дешёвое лечение по сравнению с белыми, то он в своих прогнозах системно воспроизводил заученную ситуацию несмотря на то, что положение вещей в последние годы поменялось. Не удивительно, что выдаваемый прогноз оказался некорректным. В связи с таковыми эксцессами требуется дальнейшее приложение усилий исследователей ИИ для идентификации и исправления подобных ошибок работы цифровых платформ. В будущем желательно, чтобы алгоритмы ИИ могли проходить систематическое тестирование перед внедрением, и тогда надёжность работы нейронных сетей ни у кого не вызовет сомнений [122, 128, 139, 141].



Вот некоторые часто встречающиеся ошибки девелоперов продуктовых платформ ИИ, предиктивного анализа больших данных, нейросетей и машинного обучения:

- производство продукта, исходя из имеющейся технологии, а не из определения проблем целевой аудитории;
- создание самообучающихся алгоритмов под потребности отдельно взятой категории врачей-специалистов, а не системы оказания медицинской помощи в целом;
- незнание реальной рыночной конъюнктуры и уже существующих продуктовых линеек других производителей;
- отсутствие чёткого понимания функционирования системы финансирования медицинской помощи и анализа потребностей пациентов;
- отсутствие у руководителя медицинской организации должного уровня профессиональных компетенций, в том числе, бизнес- и IT-компетенций;
- ориентация на модель «как должно быть» без предварительного формирования и отработки модели «как есть».

В дальнейшем реальная медицинская практика, скорее всего, будет включать в себя установки «человек в цикле», когда люди активно сотрудничают с системами ИИ и обеспечивают надзор. В совместных установках между ИИ и людьми медики получают помощь от него, хотя могут работать отдельно, а поступающие заключения и прогнозы впоследствии усредняются или интегрируются иным образом. Исследования в этой сфере показали, что клинические эксперты и ИИ в сочетании достигают лучших результатов, чем одни лишь эксперты. Например, учёные обнаружили, что клинические эксперты с помощью нейросети превзошли как людей, так и только ИИ при обнаружении злокачественных новообразований на рентгенограммах грудной клетки. Полезность сотрудничества человека и ИИ, очевидно, будет зависеть от специфики задачи и клинического контекста. Однако до сих пор остаются открытыми вопросы о том, как именно помощь ИИ влияет на производительность человека. В частности, установлено, что его помощь при комбинированной деятельности повышает результативность клинической экспертизы без снижения её специфичности. Интересно, что некоторые клиницисты могут получить больше пользы от помощи нейросети, чем другие; результаты исследований демонстрируют, что менее опытные врачи, например, стажёры получают больше пользы от ввода в их работу ИИ, чем их более опытные коллеги [7, 10, 18, 27, 59].

Для противодействия устоявшимся предубеждениям некоторых врачей была создана программа анализа изображений Sagimet. Проблема в том, что если алгоритмы обучения нейросети продуманы недостаточно глубоко, то она вполне способна, самостоятельно усваивая пристрастные мнения, накапливать ошибки. Это было доказано в недавних исследованиях университета Чикаго, США. Технические недоработки также могут послужить причи-



ной того, что ИИ не сможет верно интерпретировать определённые факторы, как это часто случается при моделировании развития нозологий [99, 168].

Дополнительные проблемы возникают также из-за больших размеров изображений, потому что объём памяти, требуемый нейронной сетью, может увеличиваться как со сложностью модели, так и с числом пикселей во входных данных. В результате многие медицинские изображения, в особенности изображения целых комплектов слайдов, которые легко могут содержать миллиарды пикселей каждое, слишком велики, чтобы вписаться в обычную нейронную сеть. Существует множество способов решения этой проблемы (Takhchidi K., 2021). Изображения могут быть уменьшены в размере за счёт снижения детализации, или их можно разбить на несколько небольших фрагментов, хотя это будет препятствовать способности системы проводить связи между различными областями изображения. В других случаях люди могут идентифицировать интересующую область меньшего размера, например, часть изображения слайда, содержащую опухоль, и обрезать изображение перед тем, как передать его в среду ИИ. В некоторых исследованиях используются большие пользовательские модели, которые могут принимать целые медицинские изображения, но для работы таких моделей может потребоваться дорогостоящее оборудование с большим объёмом памяти. Таким образом, системы классификации медицинских изображений зачастую требуют компромиссов, чтобы сделать входные данные совместимыми с нейронными сетями [144, 154, 161].

Проблемы также возникают, когда технологические факторы приводят к систематической ошибке смещения. Это происходит, например, если платформа использует массив данных в виде коллекции изображений, поступающих только с одной камеры с фиксированными настройками. Такая платформа покажет себя менее эффективной при обработке входных данных, полученных ею из других источников. Чтобы улучшить интерпретацию, алгоритм должен проходить обучение на базе всех тех мест, откуда будут поступать данные, чтобы адаптироваться к специфическим особенностям каждого из них. Однако и к данному подходу следует относиться с осторожностью, особенно когда распределение меток в разных наборах данных сильно различается. Например, если алгоритм обучается на наборах данных из двух учреждений, одно из которых содержит только положительные случаи, а другое только отрицательные, то он может достичь высокой производительности с помощью ложных меток, не узнавая о соответствующей патологии. Поскольку обученный таким способом алгоритм классификации изображений может начать полностью основывать свои прогнозы на различиях между камерами обоих учреждений, то он, скорее всего, узнает мало что полезного о целевой патологии и не сможет толком ничего обобщить. Поэтому-то и следует крайне настороженно относиться к технологической пред-

взятости ИИ, особенно при использовании им данных из различных источников [135, 140].

Важным аспектом завоевания нашего доверия является способность нейронной сети объяснять нам, как она приходит к своим выводам. Поскольку многие системы ИИ в настоящее время функционируют в формате чёрного ящика, объяснение результатов их работы представляет собой острую проблему. Конечно, известен ряд методов для понимания механизма действия ИИ, например, метод значимости, когда выделяются области изображения, которые могут в наибольшей степени способствовать прогнозированию. К сожалению, эти методы никак нельзя признать абсолютно надёжными. Потребуются дальнейшие исследования в области интерпретации процессов принятия решений ИИ и количественной оценки их надёжности. Когда медицинская цифровая платформа выдвигает новые идеи, выходящие за рамки текущих научных знаний, полная ясность всех шагов функционирования алгоритма особенно необходима. Только так мы сможем достойно оценить эти новые идеи, а значит и лучше понять реальные патофизиологические механизмы, лежащие в основе того или иного заболевания [47, 126, 136].

Дальнейшая интеграция ИИ в систему здравоохранения и социальный институт медицины будет связана с возможностью конкретного решения многих из вышеперечисленных проблем. Хотя существует надежда, что ИИ снизит медицинские расходы, устройства, необходимые для получения входных данных для систем ИИ, наверняка окажутся достаточно дорогими. В частности, оборудование, необходимое для получения изображений целых препаратов, является весьма дорогостоящим и потому недоступно для многих медицинских учреждений, что затрудняет как сбор данных, так и развёртывание глубоко обученных нейросетей для патологии [22, 37].

Настоящая монография предлагает обзор приложений медицинского искусственного интеллекта, а также анализ его потенциала для будущих преобразований в системе здравоохранения. Авторы постарались определить, дифференцировать и оценить текущие и потенциальные угрозы применения ИИ в медицине. Одновременно нами рассмотрены основные клинические, социально-этические и нормативные аспекты использования медицинского ИИ. Наконец, в монографии дается описание некоторых предвзятостей и заблуждений медицинских работников в определении сильных и слабых сторон нейросетевых платформ, а также предложены варианты стратегий, направленных на оптимизацию работы с ними.

## **1. ПРИЛОЖЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЗДРАВООХРАНЕНИИ**

Генерируемый медицинской наукой информационный диапазон имеет неуклонную тенденцию к расширению. Он будет и впредь продолжать расти ускоренными темпами как по объему, так и по разнообразию по мере того, как происходит совершенствование компьютерных технологий. Поэтому не вызывает никаких сомнений, что интеграция IT-платформ на основе искусственного интеллекта в медицинскую практику приведет к существенным изменениям отечественного здравоохранения.

В настоящее время внедрение нейросетевых платформ можно наблюдать практически во всех областях медицины: от неотложной помощи, до диагностики и лечения редких заболеваний, биомедицинских исследований, общественного здравоохранения и проч. Появились описания использования ИИ при организации первичной медико-санитарной помощи в поликлинике (Романов С.В. с соавт, 2024) [27]. Ожидается, что многие аспекты управления, связанные с администрированием медицинских учреждений, например, повышение эффективности производственной деятельности учреждения, контроль качества оказания медицинской помощи, работа с обращениями и жалобами граждан, также выиграют от новых инструментов, основанных на искусственном интеллекте. Инструменты ИИ в здравоохранении удобно классифицировать по группам стейкхолдеров: 1) пациенты и их представители; 2) врачи и лица, осуществляющие уход; 3) сотрудники администрации медицинских организаций; 4) чиновники из аппарата управления общественным здравоохранением муниципальных образований и территорий; 5) политики разных уровней. Классификацию биомедицинских инструментов ИИ можно основывать на условиях их использования: 1) клинические условия (специализированные стационары, поликлиники, станции скорой помощи); 2) лечебно-диагностическая деятельность (лабораторный блок, лучевая диагностика, аптека, радиология и т.д.); и 3) управление медицинским учреждением. Также целесообразно дифференцировать приложения ИИ по назначению:

1. Клинические платформы.
2. Исследовательские платформы.
3. Административно-управленческие платформы.
4. Платформы, обслуживающие нужды общественного здравоохранения (статистика и т.п.).

### **Клинические нейросетевые платформы**

Потенциал применения ИИ в клинических условиях огромен и варьируется от автоматизации диагностических процессов, принятия решений в отношении тактики лечения пациента до

клинических исследований и испытаний (Глизница П.В., 2022). Данные, необходимые для диагностики и лечения, поступают из многих источников, включая записи в историях болезни, лабораторные анализы, информацию аптеки, результаты лучевой диагностики, геномную информацию и т.д. ИИ играет важную роль в автоматизация анализа изображений (в радиологии, офтальмологии, дерматологии, патоморфологии и проч.), обработки сигналов различных датчиков (анализ электрокардиограмм, аудиограмм, электроэнцефалограмм и проч.). Помимо интерпретации изображений и результатов лабораторных тестов, ИИ можно использовать для интеграции и сопоставления результатов с другими клиническими данными для облегчения лечебно-диагностического процесса. Имеется множество впечатляющих примеров применения инструментов ИИ в клинике, в том числе, таких как офтальмология, нефрология и персонализированная медицина [14, 42, 50, 152].

Радиология также входит в число медицинских направлений, где в последние годы наблюдаются настоящие прорывы в применении нейронных сетей. Технологии ИИ для визуализации вовсю помогают в рентген диагностике и количественной оценке медицинских изображений. Например, использование моделей глубокой обучения позволяет автоматически локализовать и очерчивать границы анатомических структур и их поражений практически без участия специалиста. Эти инструменты ИИ также могут расставлять приоритеты и отслеживать результаты, требующие раннего внимания, и позволяют рентгенологам сосредоточиться на изображениях, которые, скорее всего, являются аномальными. Хорошим примером таких инструментов сегментации медицинских изображений является приложение для визуализации сердечно-сосудистой системы cvi42, продвигаемое канадской компанией Circle CVI и принятое более чем в 40 странах мира [54, 56].

Радиомика является классическим примером инновационных методов обработки изображений, в котором ИИ сегодня играет ключевую роль. Она нацелена - на извлечение количественной информации, т.е. так называемых радиомических характеристик изображений, которые отражают характеристики тканей и поражений, такие как гетерогенность и форма, и которые необходимы для принятия врачебных решений в наиболее проблемных случаях отдельно или в сочетании с данными демографических, гистологических, геномных или протеомных исследований.

Термин *цифровая патоморфология* изначально использовался для обозначения процесса оцифровки слайдов с помощью сканирования. Теперь это также относится к подходам на основе ИИ для обнаружения и анализа оцифрованных изображений. Хотя использование стандартизированных руководств, безусловно, способствует точности диагностики, но сама по себе точность гистопатологического анализа оказывается ограниченной из-за его



субъективного характера и частого расхождения интерпретаций независимых экспертов [64, 76, 77].

В этой связи ИИ может способствовать решению ряда проблем, с которыми сталкиваются онкологи и патологи, включая вариативность интерпретаций специалистов. Нейросеть вполне может обеспечить диагностическую точность, аналогичный уровню точности врача-патологоанатома, и, что более важно, может улучшить их интегральный результат при совместном использовании. В цифровой патоморфологии ИИ широко применяется для решения различных задач обработки и классификации изображений. К ним относятся и ориентированные на распознавание объектов задачи низкого уровня, такие как обнаружение патологии, и задачи высокого уровня, такие как прогнозирование течения заболевания, оценка его тяжести и реакции больного на проводимую терапию [88, 95].

Нейросети могут быть полезны учреждениям здравоохранения, оказывающим неотложную медицинскую помощь. ИИ является универсальным инструментом анализа данных анамнеза заболевания и используется на разных этапах ведения пациентов, в частности, для уточнения приоритета тех или иных тактических подходов к лечению во время сортировки пациентов. В настоящее время врачи отделений неотложной помощи проводят диагностику в условиях недостатка информации и времени для её поиска, поэтому платформы ИИ существенно расширяют возможности по оптимальному распределению ресурсов и определению показаний для проведения тех или иных вмешательств. При этом у врачей всё же сохраняется некоторая обеспокоенность в отношении использования ИИ при оказании помощи пациентам, учитывая отсутствие твёрдой уверенности в его абсолютной безопасности. Согласно обзору 150 научных публикаций, выполненном в 2019 г. Kirubakaran A. с соавт., применение нейросетей в неотложной медицине в большинстве случаев связано с их прогностическими возможностями, помощью в диагностике и сортировке пациентов, а также с их содействием в организационном планировании и административном управлении [97, 109, 121].

В практике неотложной хирургии врачам зачастую приходится принимать решения в условиях ограниченного времени и неопределённости относительно диагноза и прогнозируемой реакции пациента на предполагаемое лечение. Таковая неопределённость связана с отсутствием достоверных анамнестических данных о заболевании пациента и/или отсутствием результатов диагностических тестов, т.е. достаточно убедительных сведений, которые могли бы служить веским основанием для принятия верного решения. В данных условиях врачам приходится полагаться лишь на когнитивные методы, собственный опыт и интуицию, что не исключает предвзятости, ошибок и случайного нанесения ущерба здоровью больного. Не секрет, что порой традиционные методы принятия врачебных решений оказываются недостаточно оправданными и приводят к негативным результатам



предпринятых хирургических вмешательств (Loftus T.J., 2020). Указанные проблемы вполне преодолимы с помощью моделей искусственного интеллекта. Фактически, инструменты ИИ помогают в нахождении информации о факторах риска для конкретного пациента, истории его болезни и т.д., что, без сомнения, обеспечивает правильную тактику ведения пациента и выполнения необходимых хирургических операций (Hashimoto D.A., et al., 2019) [52, 82, 107].

Прогнозирование нейросетями клинических рисков основывается на оценке вероятностей для широкого спектра вариантов течения заболевания, вплоть до летального исхода. Процедура включает идентификацию лиц с определенными состояниями, а также классификацию по стадии, тяжести и иным характеристикам заболевания. В своих дальнейших действиях врачи могут руководствоваться полученными от ИИ рекомендациями.

Модели прогнозирования риска уже давно используются в здравоохранении. Однако до сих пор они основывались на регрессионном анализе и подмножествах доступных клинических данных, что, естественно, приводило к ограниченной точности прогнозирования и делало их мало востребованными в клинике. Технологии ИИ, работающие с большими хранилищами данных, убедительно продемонстрировали более высокую эффективность по сравнению с традиционными статистическими методами прогнозирования (Jamthikar, et al., 2020) [91, 96].

Корректирующие мероприятия по лечению пациента, иначе называемые *адаптивными вмешательствами точно в срок*, представляют собой подстраиваемые к постоянно меняющемуся состоянию больного схемы ухода и терапии. Они обеспечивают регулировку частоты, продолжительности приема и дозировки препаратов на протяжении всего курса лечения. ИИ улучшает проведение корректирующих мероприятий за счет возможности прямого участия пациента и/или посредством пассивного сбора с помощью специальных датчиков данных о функционировании организма больного. Использование нейросетей для включения мнения пациента в схему лечения называется *экологической моментальной оценкой*; она помогает больному принимать участие в слежении за лечебным процессом (De Vries, et al., 2021). Экологическая моментальная оценка важна при расстройствах, связанных с употреблением психоактивных веществ — она позволяет коррелировать проявление патологической тяги к препарату с эксцессами неадекватного поведения. Пассивный сбор данных осуществляется путем записи траектории передвижений пациента с помощью глобальной системы позиционирования GPS и беспроводных локальных сетей Wi-Fi. Возможность собирать пространственную и временную информацию о поведенческих нарушениях субъекта делает данные инструменты весьма востребованными. При совмещении данных от специальных датчиков, измеряющих артериальное давление, частоту сердечных сокращений, температуру или концентрацию тех

или иных веществ в крови, с пространственными и временными характеристиками вырисовывается подробный профиль поведения пациента, который позволяет судить по минимальным физиологическим реакциям организма о приближающемся приступе патологического влечения (Quaglio, et al., 2019) [66, 132, 133].

Платформы ИИ способны играть выдающуюся роль в повышении самостоятельности страдающих хроническими заболеваниями пожилых людей. Самоконтроль пожилому человеку необходим при приеме лекарств, соблюдении диеты, обращении с носимыми медицинскими устройствами. Домашний мониторинг обеспечивает независимость престарелых людей, отслеживая их физическое положение в случае внезапного падения. Установка на смартфоне пациента специальных мобильных приложений позволяют ему более активно участвовать в процессе лечения, облегчая взаимодействие с системой здравоохранения (Sapci, et al., 2019) [141, 156].

Тем не менее, концепция «умного дома» не может быть принята безоговорочно, поскольку имеет ряд серьезных недостатков, а именно:

- в корне меняется образ жизни пользователя;
- технологии умного дома несовершенны и сложны в эксплуатации;
- отсутствует совместимость между подсистемами;
- нет гарантии конфиденциальности и безопасности функционирования. Поэтому несмотря на выдающиеся достижения в программировании и аппаратном обеспечении, полноценное внедрение инновационных технологий «умного дома» по-прежнему остается под вопросом, что побуждает научное сообщество к поиску новых стратегий и подходов (Azzi, et al., 2020) [48, 73].

Наиболее перспективным применением ИИ является автоматизация обработки данных визуализации сердца, которая необходима для оценки структуры и функции сердечной мышцы в кардиологии. Такие методики как УЗИ сердца, компьютерная и магнитно-резонансная томография сердечно-сосудистой системы, предоставляют сложные пространственно-временные данные, обработка которых кардиологами требует высокой квалификации и занимает много времени. Доступность методов обработки изображений сердца на основе ИИ произвела революцию в клинической практике, позволив кардиологам проводить более быструю и качественную оценку состояния пациентов (Lopez-Jimenez, et al., 2020). Модели машинного обучения в первую очередь призваны улучшить диагностические возможности эхокардиографии, которая является преобладающим методом визуализации сердца, но по-прежнему в значительной степени зависит от человеческого опыта. Ожидается, что создание более точного оборудования со встроенным ИИ позволит выявлять ранее нераспознаваемые объекты визуализации, что упростит диагностику сердечно-сосудистых заболеваний, минимизировав ограничения, связанные с интерпретацией данных человеком [108].

Это уже имеет место в электрокардиографии, где созданные с использованием больших наборов цифровых данных модели ИИ типа сверточных нейросетей глубокого обучения научились выявлять такие заболевания, как бессимптомная дисфункция левого желудочка, тихая фибрилляция предсердий, а также фенотипические особенности индивида, включая пол, возраст и расовую принадлежность (Adedinsewo, et al, 2020; Noseworthy P.A., et al, 2020) [42, 81].

Кроме того, ИИ широко используется в радионуклидной диагностике патологии сердечно-сосудистой системы для неинвазивной оценки миокардиального кровотока. Модели машинного обучения применяются при *однофотонной эмиссионной компьютерной томографии (ОФЭКТ)* single-photon emission computed tomography (SPECT) и визуализация перфузии миокарда myocardial perfusion imaging (MPI), чтобы в конечном итоге улучшить выявление и прогноз ишемической болезни сердца. Установлено, что расчёт показателя 10-летнего риска возникновения сердечно-сосудистого заболевания будет более точным, если использовать алгоритмы машинного обучения, способные экстраполировать информацию и выявлять неявные закономерности в клинических данных. Хотя кардиология, по-видимому, находится на переднем крае применения медицинского искусственного интеллекта, она всегда будет зависеть от опыта конкретного специалиста. Поэтому для врача важно активно участвовать в развитии этого направления, чтобы методы обработки изображений полностью раскрыли свой потенциал и даже произвели настоящую революцию в терапии сердечно-сосудистых заболеваний (Quer, et al., 2021) [133, 165].

Хотя о применении ИИ в нефрологии сообщается достаточно редко, его потенциал все чаще признается клиницистами благодаря многообещающим достижениям последнего десятилетия. Новая модель глубокого обучения для ультразвуковой визуализации почек позволяет неинвазивно проводить дифференциальную диагностику почечной патологии. Глубоко обученная нейросеть уже способна классифицировать гистологические образцы и аннотировать результаты биопсии почек. С помощью ИИ ученым удалось усовершенствовать методику раннего обнаружения острого повреждения почек, эпизод которого теперь в стационаре можно спрогнозировать за 48 часов до начала клинических проявлений. Кроме того, алгоритм *встроенной аналитики интраоперационных данных* Intraoperative Data Embedded Analytics (IDEA) был обучен прогнозировать риск развития послеоперационного почечного повреждения путем слияния до и послеоперационных физиологических показателей (Adhikari, et al., 2019). ИИ также имеет потенциал в компьютерной диагностике рака почки. Поскольку алгоритмы становятся все более надежными и общепотребительными, они все лучше и лучше выявляют почечные новообразования, хорошо различая доброкачественные и злокачественные опухоли. Внедрение моделей ИИ в нефрологии облегчает прогнозирование и тем самым значительно

укрепляет авторитет персонализированной медицины (Giulietti, et al., 2021) [43, 76].

Модели ML широко используются для облегчения диагностики множества типов заболеваний печени, большинство из которых опасны для жизни. Интерес в первую очередь представляет раннее выявление неалкогольной жировой дистрофии печени, поскольку у большинства пациентов данное заболевание протекает бессимптомно вплоть до развития цирроза печени. Недавно разработанная нейронная сеть искусственного интеллекта демонстрирует высокую точность диагностики неалкогольной жировой дистрофии печени с точностью 97,2% (Okanoue, et al., 2021). Важно отметить, что одна и та же модель способна различать пациентов с неалкогольной жировой дистрофией печени и пациентов с более запущенной фазой заболевания в виде стеатогепатита. Также были разработаны прогностические модели для оценки тяжести и прогноза хронического вирусного гепатита, а также острой или хронической печеночной недостаточности. Несмотря на значительный прогресс в области гепатологии, ряд состояний остается недостаточно хорошо изученными в отношении применения ИИ, например, заболевания печени, связанные со злоупотреблением алкоголем, а также генетические/аутоиммунные заболевания печени (Ahn, et al., 2021) [44, 124].

У платформ ИИ имеется потенциал в вопросе оказания помощи людям с психическими расстройствами особенно в ситуациях, связанных с нехваткой квалифицированного персонала. Фактически, в настоящее время разрабатываются приложения по цифровому отслеживанию депрессии и настроения больных посредством того, как они взаимодействуют с клавиатурой, их речи, голоса. Уже существуют технологии распознавания лиц, создаются различные интерактивные чат-боты. Вычислительных мощностей систем искусственного интеллекта оказалось вполне достаточно для выявления сложной патофизиологии психических расстройств и, соответственно, для улучшения терапии. Машинное обучение помогает при: прогнозировании эффективности лечения антидепрессантами, вероятности суицида, диагностике депрессии или при приближении психотического эпизода у больного шизофренией. ИИ может помочь дифференцировать типы деменций, а также диагнозы с перекрывающимися клиническими проявлениями, но с разными стратегиями лечения в случае биполярной и униполярной депрессии. Изучение содержания и языковых моделей аккаунтов психически нездоровых людей в социальных сетях создало новые возможности для прогнозной психиатрической диагностики. Психические состояния стали наблюдаемыми в онлайн-пространстве, а информация из социальных сетей, проанализированная с помощью машинного обучения, уже используется для установки диагноза и фиксации рецидива заболевания (Lee, et al., 2021) [102, 103].



## Применение нейросетей в биомедицинских исследованиях

По всей видимости, современные нейросетевые технологии завоевали гораздо большую популярность при проведении биомедицинских исследований по сравнению с клиническими приложениями. Сегодня алгоритмы машинного обучения широко используются для ранжирования результатов поиска новой информации при накоплении медицинских знаний, эти алгоритмы постоянно самосовершенствуются, неустанно отслеживая и фиксируя направление поискового поведения пользователей. Технологии искусственного интеллекта, реализованные в поисковой платформе биомедицинской литературы PubMed для оптимизации функции поиска, включают алгоритмы машинного обучения и обработки естественного языка, которые обучаются на закономерностях, обнаруженных в действиях пользователей. Например, такой алгоритм поиска как Best Match использует опыт пользователей PubMed и передовую технологию машинного обучения в качестве альтернативы традиционному способу сортировки массива данных по дате публикации. Best Match способен ранжировать прошлые поисковые запросы пользователей по десяткам параметров, наиболее важными из которых являются: тип и релевантность статьи, дата публикации, количество просмотров, благодаря чему результативность актуального поиска многократно возрастает. Таким образом, автоматически обобщая извлекаемую информацию, глубоко обученные нейросети превращают статично повествовательные статьи в ценные клинические данные, которые могут быть применены в лечении конкретного пациента (Fiorini, et al., 2018) [9, 15, 70].

Разработчики лекарственных препаратов широко применяют методы машинного обучения для извлечения информации о новых веществах из больших баз данных о химических соединениях. Центральное место здесь занимает помощь искусственного интеллекта в деле инновационного моделирования, основанного на больших наборах данных. В результате ученым удастся добиваться большей эффективности и безопасности вновь создаваемых лекарственных средств. В частности, глубоко обученная нейросеть успешно используется для прогнозирования антимикробной активности тех или иных молекулярных структур. Алгоритм проверил более одного миллиарда молекул и, виртуально протестировав более 10<sup>7</sup> миллионов комбинаций, рекомендовал 8 соединений, которые потенциально обладают антибактериальными свойствами, но структурно весьма далеки от уже известных антибиотиков (Stokes, et al., 2020). По сравнению с традиционными испытаниями на животных, тестирование *in vitro* & *in silico* на ранних стадиях исследований может снизить стоимость разработки и количество выбраковок неподходящих вариантов. ИИ имеет большой потенциал как метод оценки соединений по их биологическим свойствам и токсичности. Существующие модели искусственного



интеллекта, основанные на методе QSAR (quantitative structure-activity relationship), т.е. количественного соотношения структуры и активности, могут использоваться при прогнозировании биологической активности новых веществ, предназначенных для различного применения. Из-за того, что модели QSAR предсказания новых соединений имели ряд ограничений за последнее десятилетие были предприняты усилия по их совершенствованию на основе методов высокопроизводительного скрининга HTS (high-throughput screening). HTS представляет собой процесс, в ходе которого с использованием стандартизированного протокола могут быть проверены тысячи и миллионы молекулярных соединений. Благодаря совмещению HTS и комбинаторного химического синтеза в современных алгоритмах ИИ мы можем оперировать огромными объемами биологических данных (Zhu, et al., 2020) [151, 169].

Рандомизированные контролируемые исследования RCT (randomised controlled trials) являются наиболее надежным методом оценки рисков и преимуществ любого медицинского вмешательства. Однако их проведение не всегда осуществимо. Неудачи RCT связаны с трудностями в подборе пациентов, неадекватной рандомизацией, недостаточным размером выборки и ошибками в определении ключевых конечных точек. Передовые платформы ИИ помогают лучше отбирать участников, точнее фиксировать и оценивать конечные точки, тем самым обеспечивая более эффективное выполнение и большую статистическую мощность исследования, чем можно ожидать от традиционного RCT (Lee, et al., 2020) [102, 104, 157].

Персонализированная медицина во многом опирается на научное понимание того, как уникальные характеристики отдельного пациента, такие как молекулярный и генетический профиль, делают его уязвимым к тому или иному заболеванию и чувствительным к рекомендованной терапии. Вклад множества генов в развитие патологии человека уже известен, поэтому генетические особенности больных начали активно использоваться для прогноза индивидуальной реакции на лечение (Strianese, et al., 2020). Сегодня за счет включения ряда дополнительных источников и индивидуальных клинических характеристик в изначальную концепцию персонализированной медицины создана новая концепция «расширенной персонализированной медицины». Этими источниками послужили клинические, демографические, социальные данные, физиологические параметры, касающиеся образа жизни пациента (часы сна, физическая активность, привычки питания и т. д.), а также условия окружающей среды. Инструменты ИИ способствуют прогрессу персонализированной медицины за счет взвешенной оценки клинической пользы различных методов исследования, учета множества типов данных, прогнозирования эффективности создания и применения целевых лекарственных препаратов, компьютерного моделирования метаболических процессов в организме, идентификации закономерностей и разнообразия популяционной генетики. Таким образом, переход к расширенной

концепции персонализированной медицины, основанной на поддержке ИИ, будет иметь далеко идущие последствия для населения, врачей и фармацевтической промышленности (Bonniolo, et al., 2021) [11, 53, 152].

### **Применение ИИ в сфере общественного здравоохранения**

Одно из определений общественного здравоохранения, данное D. Wanless в 2004 г. таково: «Это наука и искусство предотвращения болезней, продления жизни и укрепления здоровья посредством организованных усилий и осознанного выбора общества, организаций, государственных и частных, сообществ и отдельных лиц». В настоящее время работы с соответствующими платформами искусственного интеллекта проводятся в целом ряде направлений развития общественного здравоохранения. ИИ может помочь на основании демографических и географических данных локализовать регионы, где распространены те или иные заболевания или высоко рисковое девиантное поведение отдельных групп населения. Спектр решений искусственного интеллекта в области эпидмониторинга постоянно расширяется. Цифровой эпидемиологический надзор подразумевает интеграцию сведений о событиях из новостной ленты, социальных сетей, от датчиков мобильных устройств, цифровых следов, микробиологических лабораторий, клинических отчетов, необходимых для анализа информационного потока и проверки на наличие угроз. Такой подход был реализован при создании систем раннего предупреждения о неблагоприятных последствиях употребления наркотиков и загрязнения воздуха. Благодаря данным, генерируемым датчиками и роботами, ИИ уже прочно вошел в арсенал технологий по охране окружающей среды и гигиены труда. Важным преимуществом ИИ является возможность установления и поддержания контакта с населением с помощью различных автоматизированных, легко масштабируемых методов, таких как обмен текстовыми сообщениями и порталы для пациентов. Платформы ИИ нацелены на повышение эффективности работы медицинских учреждений, улучшение распределения системных ресурсов здравоохранения, в частности, для экстренного оказания помощи при дорожно-транспортных происшествиях [3, 4, 26, 161].

Администрирование системы здравоохранения – это самый сложный комплекс трудоемких процессов с широким кругом участников и стейкхолдеров, включая пациентов, медицинских работников, вспомогательный персонал медицинских учреждений и организаций, инфраструктуру, материально-технические средства, например, средства визуализации, лабораторное оборудование, аптеку и т.д. Многочисленные исследования, проводимые в учреждениях здравоохранения, выявили несколько особенно проблемных аспектов в этой многокомпонентной административной среде, к которым следует отнести: время, затрачиваемое на подготовку и осуществление информационных транзакций, в том числе, финансового характера,

ввод данных в различные слабо интегрированные с корпоративной информационной системой программы и приложения, обработку и интерпретацию данных, поступающих от больниц и прочих внешних источников, а также оперативную помощь пациентам в навигации по фрагментированной системе здравоохранения. По мнению большинства исследователей, не менее 50% общего количества времени и ресурсов здравоохранения расходуется на вопросы бюрократического свойства, львиную долю которых составляют проблемы, которые можно было решить намного быстрее и с гораздо меньшими затратами (Clay & Stern, 2015). ИИ может выполнять столь рутинные задачи на несколько порядков быстрее, точнее и совершенно беспристрастно. Убедительным аргументом в пользу использования административно-хозяйственных платформ ИИ является то, что просчеты в этой сфере напрямую не затрагивают жизни и здоровья человека и не так опасны, как врачебные ошибки. Однако сохраняются риск взлома и угроза конфиденциальности. Алгоритмы ИИ имеют решающее слово при управлении потоками пациентов. Например, хирургические вмешательства часто откладываются из-за отсутствия свободных коек в отделениях постоперационного выживания. Эта проблема связана с чисто административными задержками при переводе больных в клинические отделения или при выписке из стационара [60, 65].

Кодификация – это сложный и трудоемкий процесс извлечения информации из клинических записей и ее шифрования с использованием таких классификаций, как, например, Международная классификация болезней (МКБ). Точность кодификации очень важна для администрирования, последующего возмещения расходов медицинского учреждения и проведения научных исследований. Хотя компьютерное кодирование существует уже более десяти лет, ИИ может повысить точность и существенно упростить выполнение данной процедуры. Планирование является ещё одним хорошим примером того, как ИИ может помочь в организации производственной деятельности медицинского учреждения. Алгоритмы, основанные на больших данных из историй болезни, могут предсказывать, кто из запланированных пациентов не явится на прием, что позволяет лечащим врачам принимать упреждающие меры. Помимо целевых напоминаний пациенту ИИ при необходимости может удовлетворять некоторые его потребности и запросы. Нейросеть также может быть обучена обнаруживать мошеннические действия и схемы в сфере здравоохранения, допустим, использование кода более дорогостоящей медицинской услуги вместо реально оказанной [51, 52].

Беспрепятственное продвижение пациентов по всем этапам оказания медицинской услуги – вот что характеризует оптимально организованный поток пациентов. Примечательно, отсутствие задержек на маршруте движения – главный критерий качества и доступности медицинской помощи, который напрямую влияет на конечную удовлетворенность пациентов и их представителей

(Власов Я.В., 2024). Убедительно доказано, что плохо налаженная производственная логистика учреждения здравоохранения крайне отрицательно влияет на пациентов, персонал и конечные результаты оказания медицинской помощи (Бердугин В.А., 2020). Инновационные технологические решения на основе ИИ будут весьма востребованы для управления потоковыми процессами поликлиник и стационаров, поскольку неравномерная мощность входящего потока пациентов, будучи неопределенной переменной, является причиной сильной головной боли многих руководителей приёмных отделений и больниц неотложной помощи. Заблаговременный прогноз количества поступающих пациентов позволяет правильно спланировать производственную нагрузку и обеспечивает своевременность принятия соответствующих решений (Dawoodbhoy, et al., 2021). Аудит в сфере здравоохранения представлен государственным, ведомственным и внутренним видами контроля качества и эффективности медицинской помощи. Процедура аудита предоставляет нам как количественную информацию о текущем положении дел в учреждении, так и рекомендации того, как можно улучшить результаты диагностики и лечения [8, 12, 65].

### **Риски применения ИИ в здравоохранении**

В 1970 г. W. Schwartz писал, что достижения в области информатики рано или поздно приведут к созданию искусственного интеллекта, способного заменить мышление врача. Несмотря на впечатляющие примеры нейросетевых решений в области медицины, данное предсказание ещё очень далеко от своего воплощения. Нынешние результаты применения ИИ в здравоохранении не столь надежны, как предполагалось ранее, поэтому его будущее пока видится достаточно расплывчатым [147].

Отдельные учёные думают, что потенциал медицинского ИИ сильно переоценен, ведь слишком мало убедительных доказательств реального улучшения состояния больных, лечившихся исключительно под его патронажем. Более того, многие эксперты выражают искреннюю обеспокоенность по поводу очевидных рисков массового применения медицинского ИИ, грозящих неблагоприятными клиническими и социально-экономическими последствиями (Morley & Floridi, 2020; Manne & Kantheti, 2021). В специальной литературе есть упоминания о множестве неприятных сторон использования медицинского ИИ, которые можно условно распределить по категориям:

- ущерб, наносимый пациентам, вследствие неверных решений ИИ;
- неграмотное использование инструментов медицинского искусственного интеллекта;
- проявление предвзятости и бескомпромиссности ИИ;
- непрозрачность решений ИИ;
- ненадежность гарантий безопасности и защиты конфиденциальности;



- сложности с обеспечением подотчетности ИИ;
- проблемы с внедрением и массовым практическим применением в организациях здравоохранения;
- дальнейшее развитие технологий ИИ может неожиданно привести к появлению сверх интеллектуальной нейронной сети, которая во всех аспектах превзойдет умственные способности даже самых гениальных людей, сможет выйти из-под контроля человека и устроить планетарный апокалипсис.

Описанные минусы и риски ИИ способны не только принести вред больным и доставить проблемы медперсоналу, они также могут подорвать доверие к алгоритмам ИИ у общества в целом. Поэтому оценка, классификация и управление рисками должны быть неотъемлемой частью процессов разработки, внедрения и эксплуатации любых медицинских нейросетевых платформ [110, 116].

Можно выделить следующие варианты случаев нанесения ущерба пациентам из-за сбоев и ошибок в работе алгоритмов ИИ:

- ложно отрицательный результат заключения ИИ в виде пропущенного диагноза опасного для жизни пациента заболевания,
- назначение ненужного лечения из-за ложно положительных результатов заключения ИИ, когда здоровые люди ошибочно признаются больными,
- проведение ненужных оперативных вмешательств или манипуляций из-за неверной диагностики и неправильной расстановки приоритетов в приемных отделениях и при оказании неотложной помощи.

При этом причинами данных ситуаций могут быть: недостаточный объем, артефакты и низкое качество обучающих ИИ данных; значительные расхождения обучающих данных с объективной реальностью; непредсказуемая вариативность окружающей обстановки по отношению к рутине клинического контекста, например, в условиях чрезвычайных ситуаций. Сглаживание негативных последствий предусматривает мероприятия: по проведению всеобъемлющих мульти фокальных исследований по идентификации дестабилизирующих влияний на работу платформы ИИ; по разработке вспомогательных нейросетевых решений для клинических нужд; по мониторингу за работой ИИ для её поступательного улучшения.

Несмотря на то, что разработчики программного обеспечения получают доступ к крупным датасетам вполне приличного качества для обучения нейросети, клиницисты отмечают 3 основных причины ошибок ИИ. Во-первых, это шум во входных данных, например, при ультразвуковом сканировании. Тут многое зависит от опыта врача, готовности пациента и общеклинического контекста, допустим, если УЗИ проводится по экстренным показаниям. Во-вторых, весьма распространенной проблемой машинного обучения является сдвиг датасета, возникающий, когда статистическое распределение клинических данных пусть даже незначительно отклоняется от исходного распределения данных в датасете, который использовался для обучения ИИ. Этот сдвиг может быть обусловлен:



различиями в группах обслуживаемого населения, использованием оборудования разных производителей, погрешностями калибровки датчиков и т.п. Недавнее исследование показало, что модели искусственного интеллекта, обученные на сканах магнитно-резонансной томографии сердца с аппаратов одной фирмы, например, General Electric теряют точность при работе с данным, полученными от МРТ-сканеров других производителей, скажем Siemens или Philips (Campello, et al., 2020) [56].

Еще с одним характерным примером сдвига датасета мы сталкиваемся, когда читаем о результатах многоцентрового исследования, в ходе которого была создана высокоточная модель искусственного интеллекта для диагностики пневмонии в двух крупных американских больницах. При тестировании модели с данными из третьей больницы было отмечено значительное падение точности, которое указывало на системную ошибку, связанную с особенностями конкретной больницы (Zech, et al., 2018). В другом примере компания DeepMind разработала модель глубокого обучения, обученную на большом наборе данных для автоматизированной диагностики заболеваний сетчатки с помощью оптической когерентной томографии. Однако выяснилось, что ИИ дал сбой при анализе изображений, полученных из аппарата, отличного от того, который использовался на этапе обучения. При этом ошибка диагностики подпрыгнула с 5,5% аж до 46%. Наконец, заключения ИИ могут быть ошибочными из-за сложности адаптации алгоритмов к неожиданным изменениям в окружающей среде и контексте, в котором они применяются. Ученые из Гарвардской медицинской школы обучили нейросеть обнаруживать тени или участки уплотнений на рентгеновских снимках грудной клетки, указывавшие на наличие рака легких. После чего им удалось выявить причины, которые легко могли привести к неправильным заключениям ИИ. Это были случаи, когда на груди пациента оставались клейкие разъемы от ЭКГ или металлические предметы типа амулета на шнурке. В этих ситуациях алгоритм вполне мог, ошибочно приняв округлые артефакты за патологическое образование, выдавать ложно положительный результат [168].

Как мы уже упоминали, существуют реальные способы минимизации рисков от ошибок ИИ и обеспечения безопасности пациентов. Прежде всего, необходимо стандартизировать методы и процедуры оценки и одобрения регулирующими органами здравоохранения новых платформ медицинского ИИ, в частности, в отношении их применимости к разным группам населения и чувствительности к шуму. Во-вторых, модели ИИ должны быть разработаны и реализованы именно в качестве вспомогательных, а вовсе не автономных инструментов для того, чтобы врачи продолжали оставаться главными действующими лицами как лечебно-диагностического процесса, так и процесса интерпретации данных, с целью своевременного обнаружения ошибок и учета изменений клинического контекста. Кроме того, будущие модели медицинского ИИ должны обязательно иметь встроенный механизм обуче-

ния на новых сценариях и прошлом опыте. Причём данный аспект потребует тщательного и многоступенчатого контроля, что, в свою очередь, безусловно, увеличит эксплуатационные затраты. Регулярное обновление программного обеспечения также потребует пристального внимания административных структур и определенных финансовых вложений с их стороны, равно как и реализация политики по обеспечению имплементации указанных требований в производственные системы учреждений здравоохранения [164].

Как и в случае с большинством медицинских технологий, при использовании медицинского ИИ существует риск его неправильного применения людьми либо из-за незнания, либо из-за банальной оплошности. Даже если сам по себе алгоритм ИИ является абсолютно точным и надежным, его использование на практике зависит от конечного потребителя: медицинского работника и/или пациента. Неумелое использование инструментов ИИ может привести к принятию ошибочных решений и причинить вред больному. Следовательно, заинтересованным лицам недостаточно просто иметь доступ к нейросетевым технологиям, им нужно хорошо понимать с каким сложнейшим инструментом они имеют дело и быть подготовленными к работе с ним.

Есть много факторов, способствующих неправильному/ошибочному использованию человеком нейронных сетей. Большинство проблем возникает из-за того, что разработчики зачастую создают их без непосредственного участия конечных пользователей и без учета мнения клинических экспертов. В результате именно на врачей и пациентов ложится бремя освоения сложной и незнакомой технологии, что, естественно, является достаточно трудным делом. Мало кого из нас можно назвать большими любителями преодолевать лишние трудности, поэтому вероятность вызвать негативные эмоции пользователей здесь будет обязательно присутствовать. А раздраженный человек более склонен допускать ошибки в работе. Эта проблема усугубляется тем фактом, что существующие программы обучения в медицинских университетах еще не полностью сочетают получение студентами клинических знаний и цифровых навыков (Романова Т.Е. с соавт., 2024), поэтому молодые врачи ощущают нехватку нужных знаний и навыков. Наша монография как раз и нацелена на то, чтобы хоть отчасти компенсировать имеющийся дефицит информации по данному вопросу [25, 155].

Опрос, проведенный в 2021 году в Австралии и Новой Зеландии с участием 632 врачей-стажеров по специальностям офтальмологии, дерматологии и онкологии, показал, что 71% респондентов считают, что ИИ в будущем будет полезен при скрининге обслуживаемого населения на наличие заболеваний и в решении рутинных задач. Однако большинство респондентов указали, что никогда не использовали приложения ИИ в своей работе в качестве врача (>80%), и только 5% считали себя относительно компетентными в этом вопросе (Scheetz, et al., 2021). В другом исследовании, проведенном в Великобритании, приняли участие

484 студента из 19 медицинских школ. Было установлено, что ни один из них не изучал основы ИИ в рамках обязательной учебной программы. Аналогичные выводы были сделаны в отношении технических знаний и навыков медиков из Европы, практикующих в разных сферах здравоохранения (Sit, et al., 2020). Похожая ситуация в отношении грамотности в области искусственного интеллекта также наблюдается у людей, которые должны стать его будущими пользователями. Исследование 2021 года, проведенное в Австралии, США, Канаде, Германии и Великобритании с участием более 6 тыс. человек, показало, что население в целом вообще плохо понимает, что такое ИИ. Молодое поколение гораздо лучше осведомлено в этом вопросе, но даже те люди, которые имели высшее образование, сообщили, что и они до конца так и не разобрались, что представляет собой ИИ и для чего он нужен (Gillespie, et al., 2021) [75, 142, 150].

Бесконтрольное распространение легкодоступных медицинских приложений ИИ также является опасным для общества фактором. Некоторые компании в коммерческих целях разработали мобильные приложения для диагностики рака кожи. Теперь кто угодно может загрузить фото участка своей кожи, которое напрямую анализируется встроенным алгоритмом. К таким приложениям относятся: Skinvision, MelApp, SkinScan, SpotMole. Хотя эти инструменты легко доступны широкой публике, имеется очень мало информации о том, как данные алгоритмы ИИ были разработаны и как была осуществлена проверка их надежности и клинической эффективности. А ведь недавнее исследование, в котором оценивалось 6 дерматологических мобильных приложений, продемонстрировало их низкую эффективность и избыточную предвзятость. Авторы пришли к выводу, что нельзя полагаться на современные приложения для смартфонов, спроектированные для обнаружения меланомы или других видов рака кожи (Freeman et al., 2020). К тому же текущая нормативная база в аспекте медицинского ИИ до конца не проработана и не гарантирует надежную защиту интересов пациента [62, 72, 130].

Появившиеся на рынке онлайн-приложения с ИИ нацелены на применение в различных областях медицины, начиная от диагностики и заканчивая мониторингом состояния больного, например, Diagnostics.ai, DDXX Doctor Ai, Symptomate, Achu Health. Хотя такие продукты удобны и представляют собой вполне неплохие решения, их свободное распространение является серьезной проблемой для здравоохранения и общества, равно как и популярные онлайн-аптеки, которые из-за легкодоступности лекарственных средств создают предпосылки для злоупотребления ими.

Поскольку производители мобильных медицинских приложений на базе ИИ получают колоссальную прибыль, этот сектор экономики будет привлекать внимание всё новых и новых компаний, в основной массе мало озабоченных соблюдением биоэтики и необходимых стандартов качества. Продавцы этих приложений открыто указывают на своих веб-сайтах, что не несут никакой

ответственности, поскольку их продукция не проходила сертификацию. В Интернете нередко встречаются такого рода заявления: «этот сайт предназначен исключительно для предоставления общей медицинской информации в образовательных целях» или «медицинская информация, представленная на данном сайте, и интерактивные советы не предназначены для использования в качестве профессиональных рекомендаций и не являются заменой врачебных консультаций». При этом большинство людей склонны следовать агрессивной рекламе, не понимая того, что за столь расплывчато прекраснодушными заявлениями, скрывается явное нежелание производителей нести юридическую ответственность за негативные последствия применения своих устройств [40, 52].

Между тем, существуют методы борьбы с неправильным использованием платформ медицинского ИИ. Главное, чтобы конечные пользователи, такие как медицинские работники, технические специалисты, и даже пациенты, активно привлекались к проектированию и разработке алгоритмов искусственного интеллекта, и чтобы их мнения обязательно учитывалось в процессе принятия окончательных решений даже по мелким вопросам. Кроме того, необходимо разрабатывать университетские образовательные программы и популяризировать курсы повышения грамотности в области медицинского ИИ среди населения, ибо, только повысив знания и навыки будущих конечных пользователей ИИ, нам удастся снизить негативное влияние пресловутого человеческого фактора. Наконец, важно, чтобы государственные контролирующие органы активно взялись регулировать сектор веб-/мобильного медицинского ИИ, чтобы граждане были полностью проинформированы и надёжно защищены от неграмотного применения малопонятных, но легко доступных нейросетевых технологий [33, 38].

Несмотря на поистине прорывные достижения в области медицинской науки в системах здравоохранения даже самых высокоразвитых стран мира по-прежнему наблюдаются социальная несправедливость и неравенство в доступности медицинского обслуживания для экономически слабых слоев населения. К основным факторам, которые способствуют этому неравенству относятся: возраст, половая и этническая принадлежность, уровень благосостояния, образование и место проживания. Хотя некоторые из этих факторов носят системный характер из-за социально-экономических различий, не менее важную роль здесь могут играть бытующие в обществе предубеждения. Например, в Соединенных Штатах есть исследования, свидетельствующие, что врачи не реагируют на жалобы чернокожих пациентов так же оперативно, как на жалобы белых. В отдельных странах мира, в той или иной степени, сохраняется еще один вариант предвзятости – дискриминация по половому признаку. Врачи этих стран обычно больше прислушиваются к сообщениям мужчин об испытываемой ими боли, чем к словам женщин в однотипной ситуации (Hoffman, et al., 2016). Обученные в таких условиях нейронные сети, закрепляя человеческие



предубеждения и предрассудки, будут только усиливать расслоение общества и способствовать неравенству в отношении доступности медицинской помощи. Вот несколько вопиющих примеров алгоритмических ошибок, которые вызвали настоящую шумиху в прессе [87].

Издание Science в 2019 году рассказало об алгоритме, используемом в США при направлении пациентов для оказания им специализированной помощи, который дискриминировал афроамериканцев. После скандала была произведена корректировка настроек и неравенство было устранено. При этом доля получающих специализированную помощь чернокожих пациентов сразу выросла с 17,7% до 46,5%. В канадском исследовании 2020 г. оценивалась степень справедливости современных алгоритмов глубокого обучения, используемых для обнаружения патологических изменений в легких при рентген диагностике. Было установлено, что наиболее часто случаи гипо диагностики наблюдались у молодых женщин до 20 лет, у цветных пациентов и у лиц, находящихся на государственной медицинской страховке для людей с низким доходом. Кроме того, лица, которым не посчастливилось, и они относились ко всем указанным категориям, например, бедные молодые женщины латиноамериканского происхождения вообще не признавались больными. Отсюда был сделан вывод, что модели ИИ, обученные на чересчур больших выборках, не способны обеспечить равенство при оказании медицинской помощи, если их эксплуатировать без соответствующей коррекции (Seyyed-Kalantari, et al., 2020). М. Ghassemi из университета Торонто заявила, что предвзятость ИИ надо рассматривать как естественное проявление реального состояния социума. Дело в том, что, когда мы тренируем алгоритмы на датасетах, сформированных отнюдь не идеальными людьми, которые в силу своей профессиональной деятельности занимаются маркировкой и комментированием данных, то мы ожидаемо сталкиваемся с целым букетом предвзятых мнений и предрассудков, которые люди неосознанно привносят в эти данные (Ghassemi, 2021). В ходе национального исследования США по скринингу легких 2002 г. было обследовано 53 тыс. курильщиков с целью ранней диагностики рака легких. Однако в выборку попало лишь 4% чернокожих участников. Так же было установлено, что алгоритмы ML для обнаружения рака кожи обучались на одной из наиболее широко известных баз данных открытого доступа International Skin Imaging Collaboration, которую никак нельзя признать образцовой, так как она содержала изображения, полученные в основном от светлокожих людей. Не случайно модели ИИ, обученные только на белокожих континентах, совершенно не годятся для диагностики меланомы у негроидов [74, 144].

Ещё одна группа искажающих датасеты предвзятостей связана с географическим местоположением. Исследователи из Стэнфордского университета при обзоре статей, описывающих обучение алгоритмов ML для ухода за пациентами, обнаружили, что в 71% исследований использовались данные только из Калифорнии,

Массачусетса и Нью-Йорка. При этом вообще не было представлено никакой информации по 34 штатам США из 50. Географическая предвзятость может стать важной мировой проблемой, поскольку в разных регионах земного шара наблюдается крайне неравномерное распределение доступности цифрового оборудования.

Отклонения функционирования медицинского ИИ также бывают связаны с предвзятостью при маркировке и оценке клинических данных. Например, из-за гендерных предпочтений у женщин более часто диагностируется такое заболевание, как депрессия, и реже определяется онкологическая патология. Крупномасштабное датское исследование, в ходе которого были проанализированы медицинские карты 7 млн. госпитализированных больных по 19 нозологиям, показало, что диагноз заболевания у подавляющего большинства женщин был установлен гораздо позже, чем у мужчин. Важно отметить, что для многих этих заболеваний, в том числе, травм, отравлений, врожденных пороков развития и инфекций, обнаруженные временные несоответствия нельзя было объяснить ни анатомическими, ни генетическими различиями. Если не искоренить фиксацию подобных расхождений при разметке данных в реестрах здравоохранения, когда определенным группам лиц под влиянием стереотипов систематически ставится ошибочный диагноз, то модели ИИ, неизбежно будут приучены считать эту несуразность нормой (Westergaard, et al., 2019) [162].

Понимая опасность, которую представляет собой алгоритмическая предвзятость ИИ, разработчики стали активнее сотрудничать с клиницистами и уделять более пристальное внимание выборке и маркировке датасетов и переменных, которые в дальнейшем планируется использовать для обучения модели. Наконец пришла осознанность, что данные должны быть репрезентативными и сбалансированными по ключевым характеристикам: пол/гендер, возраст, социально-экономические условия, этническая принадлежность, географическое местоположение. Авторы настоятельно рекомендуют привлекать в команды разработчиков помимо ученых, сведущих в биомедицинских технологиях и биоэтике, ещё экспертов в области общественного здравоохранения, социологов, и даже пациентов. Рабочая группа должна быть не только представительной, но и соционически выверенной, т.е. люди должны быть психологически совместимы друг с другом, чтобы обеспечить разнообразие стилей мышления и полноту жизненного опыта (Бердугин с соавт., 2023) [6].

Несмотря на непрерывный прогресс в области медицинского ИИ, его алгоритмы по-прежнему рассматриваются как отдельными представителями врачебного сообщества как какие-то непостижимо сложные технологии, которыми трудно управлять и которым совершенно нельзя доверяться. Алгоритм искусственного интеллекта от Google для скрининга рака молочной железы привлек внимание общественности своей прогностической точностью. В отдельных случаях его заключения оказывались точнее,

чем у врачей-рентгенологов. Однако эта информация встретила острую критику в средствах массовой информации, поскольку не содержала никаких ключевых технических подробностей о том, как был устроен данный алгоритм. В журнале Nature группа учёных, поставив под сомнение безопасность такого подхода, призвала к большей транспарентности нейросетевых платформ. Не случайно отсутствие прозрачности считается слабой стороной медицинского искусственного интеллекта, которая подрывает доверие к нему у врачей и пациентов. Концепция прозрачности ИИ тесно связана с понятиями прослеживаемости и объяснимости. Прозрачность процессов разработки и использования ИИ – это прослеживаемость, а прозрачность решений ИИ – это объяснимость [71].

Прослеживаемость считается ключевым требованием для заслуживающего доверия ИИ и относится к тщательному документированию всего процесса разработки, включая фиксацию того, как модель ИИ функционирует после развертывания в реальных условиях. В частности, транспарентность предполагает полный учет: 1) деталей модели (назначение, тип нейронной сети, её параметры, описание этапов предварительного и последующего тестирования; 2) данных обучения и проверки (процедуры и протоколы сбора, состав и маркировка данных); 3) мониторинга производительности (показатели и оценка производительности, случаи сбоев). Фактически же существующие платформы медицинского искусственного интеллекта редко обеспечивают полную прослеживаемость. Коммерческие компании предпочитают не раскрывать секретов своих алгоритмов, поэтому они зачастую представляют собой совершенно непрозрачные алгоритмы, которые невозможно беспристрастно изучить и дать независимую оценку. В то время как прослеживаемость обеспечивает прозрачность жизненного цикла разработки алгоритма ИИ, объяснимость нужна для прозрачности каждого заключения или решения модели. Европейский регламент по защите данных European Union's General Data Protection Regulation (GDPR) содержит положение *о праве на объяснение*, которое требует от разработчика программного продукта предоставления подробного объяснения процесса принятия решений, который использует алгоритм [68, 114].

Из-за отсутствия прозрачности глубокие нейронные сети называют черными ящиками. Они используют сложный для понимания человека алгоритмический комплекс, функционирование которого, как и процессы принятия решений не прослеживаемы и не очевидны. Отсутствие прозрачности затрудняет внедрение конкретных решений ИИ в практическое здравоохранение, поскольку врачи должны ясно понимать принцип работы каждой платформы, которая потенциально может быть полезна. Закрытость алгоритма означает невозможность определить причину сбоев и ошибок ИИ, а также установить виновных лиц. Существует ряд способов повысить прозрачность технологий искусственного интеллекта в здравоохранении. Прежде всего, существует необходимость в *пас-*

*порте ИИ*, который мог бы стать обязательным требованием для каждой модели ИИ при её документировании. Также существует потребность создания инструментов мониторинга производительности алгоритма после внедрения, в частности, для регистрации ошибок и для периодических проверок. Для улучшения объяснимости алгоритмов ИИ, важно, чтобы разработчики привлекали конечных пользователей и интересантов с самого начала разработки. Регулирующие органы должны играть ведущую роль, рассматривая прослеживаемость и объяснимость алгоритма ИИ в качестве базовых предпосылок предстоящей сертификации.

Потенциальные риски моделей ИИ, связанные с ненадёжностью их систем безопасности при обеспечении конфиденциальности и защиты данных пациентов, угрожают серьезными неприятностями, таким как раскрытие и использование личных данных граждан для неблагоприятных целей. Эти риски актуализируются, когда происходит обмен приватными медицинскими данными при предоставлении пациенту необходимой информации для принятия им самостоятельного решения, например, информированного согласия на операцию. Информированное согласие является важным и неотъемлемым правом граждан в сфере охраны здоровья, которое связано с такими юридическими аспектами, как защита от вреда, уважение автономии и прав собственности, обеспечение конфиденциальности [20, 21, 67].

Распространение непрозрачных алгоритмов ИИ, ведущее к усложнению процедур оформления информированного согласия ограничивает степень автономии и диапазон равноправного сотрудничества врачей и пациентов, которым становится все труднее отслеживать ход принятия решений и манипуляции с использованием их личных данных, а также отвергать условия, которые их не устраивают. Проблематика информированного согласия проступает особенно явно при анализе больших медицинских данных на основе цифровых платформ, который показывает, что обычный пациент даже не представляет себе уровень сложности технологий, которые задействованы в работе с информацией, содержащей, в том числе, и его личные данные. Вопиющим примером нарушения фундаментальных прав граждан явился случай, когда записи 1,6 млн. британских подданных без их согласия были переданы из фонда Royal Free NHS Foundation Trust в компанию DeepMind, принадлежащую Google и занимающуюся разработкой приложения для диагностики почечной патологии с помощью ИИ (Gerke et al., 2020) [73].

Использование медицинского ИИ также угрожает безопасности хранения данных, в результате чего личная информация может попасть в открытый доступ, что нарушает право граждан на неприкосновенность частной жизни, подвергает их риску цифрового мошенничества и других видов кибератак. В 2020 г. в занимающейся вопросами ИИ компании Sense AI произошла утечка данных, в результате которой были раскрыты высоко конфиденциальные данные



более чем 2,5 млн пострадавших в автокатастрофах американцев. Данное происшествие прямо свидетельствует о вполне реальном риске нарушения конфиденциальности, которому подвергаются пациенты в эпоху господства нейросетевых технологий.

Дополнительной проблемой также является перепрофилирование или ползучесть данных, когда медицинские данные перепрофилируются для целей, не связанных со здравоохранением. Хотя перепрофилирование возможно и внутри самой системы здравоохранения. Например, данные из электронных медицинских записей или реестров страховых медицинских организаций начинают использоваться для создания новых фармпрепаратов, клинических испытаний, маркетинга медицинских услуг и проч.

Помимо проблем, связанных с конфиденциальностью и безопасностью данных, нейросети особенно уязвимы для кибератак, результаты которых могут быть от обременительных до фатальных, в зависимости от ситуации. Известен случай смерти пациента в процессе его перевода в другую больницу из подвергшейся кибератаке университетской клиники Дюссельдорфа, которая привела к сбою её компьютерной системы. Хотя невозможно доказательно утверждать, что смерть была вызвана именно кибератакой, поскольку больной и так находился в критическом состоянии, тем не менее, данный случай ярко демонстрирует фатальные последствия уязвимости ИИ. Факты упрямо говорят о том, что технологическое несовершенство нейросетевых платформ опасно для здоровья пациентов. Так шведская компания по разработке программного обеспечения для онкологии Elekta подверглась атаке с помощью программы-вымогателя, которая затронула 170 клиник США и привела как к задержке лечения онкологических больных по всей стране, так и к раскрытию конфиденциальных данных пациентов (Mulcahy, 2021) [117].

Более того, персональные медицинские устройства особо уязвимы для атак. Например, управляемые ИИ инсулиновые помпы больных диабетом несложно взломать и, дистанционно манипулируя подачей инсулина в организм, спровоцировать летальный исход. Эти факты прямо говорят нам о том, как алгоритмическая безопасность, или ее отсутствие, влияет на жизнь и здоровье современного человека. Вот почему всем стейкхолдерам необходимо уделять ей самое пристальное внимание и постоянно контролировать. Крайне важно повысить осведомленность и грамотность медицинских работников в отношении рисков конфиденциальности и безопасности, а также информированного согласия и киберзащиты. Нормативная база должна быть отрегулирована таким образом, чтобы обеспечить прозрачность и подотчетность на всех уровнях, а также профилактику утечки и нецелевого использования данных. Следует продвигать инновационные подходы к ИИ, чтобы использовать возможности больших клинических баз данных для безрисковой передачи и обмена датасетами. В этом направлении нужно наращивать усилия, неуклонно повышая

безопасность облачных систем и ограждая нейросетевые алгоритмы от кибератак.

В последнее время термин *ответственность/подотчетность алгоритма* стал довольно употребительным в сообществах разработчиков, озабоченных правовыми аспектами использования алгоритмов медицинского ИИ. На первый взгляд может показаться, что данный термин как-то связан с привлечением к юридической ответственности самого алгоритма, хотя, на самом деле, все не так. Он акцентирует лишь тот факт, что нейросеть воспитывается благодаря сочетанию машинного обучения и человеческого замысла. А ошибки и сбои алгоритмов происходят от неправильных и даже преступных действия людей, разрабатывающих, внедряющих или использующих платформы ИИ. Ничего не может быть глупее, чем предъявлять компьютерной программе претензии в безответственности и аморальности. Подотчетность особенно важна для медицинского ИИ, поскольку она будет способствовать его надежности, социальной приемлемости, а также приятию и признанию в медицинском сообществе. За все промахи и ошибки искусственного интеллекта так или иначе отвечает врач, который его использует, а вовсе не компания-разработчик, поэтому медики не рвутся применять мало апробированные алгоритмы в своей повседневной практике. Пациенты также быстро утратят доверие к ИИ, если им покажется, что никто из разработчиков или использующих инструментарий ИИ операторов не собирается нести ответственность за возможный вред. Отсюда возникает потребность в правовых механизмах и регулирующих структурах для обеспечения адекватной подотчетности медицинских алгоритмов, уполномоченных в вопросах управления рекламациями, компенсациями и санкциями. К сожалению, до сих пор проблема определения ответственности алгоритма из-за неопределенности правового приоритета остается до конца не решённой.

Применение действующего законодательства и принципов ответственности к новым приложениям ИИ в здравоохранении предполагает:

- распределение ответственности между многочисленными стейкхолдерами, участвующими в разработке, внедрении и использовании медицинского ИИ;
- точную установку ведущей причины любой ошибки медицинского ИИ, т.е. связана ли она с обучающими данными, с неправильным использованием алгоритма или искаженной интерпретацией его решений;
- стандартизацию систем управления и унификацию этических и правовых норм в отрасли медицинского ИИ.

Исторически в центре проблем межгрупповых отношений в здравоохранении всегда находились вопросы профессиональных ошибок и врачебной халатности. Вплетение в существовавший клубок трудно разрешаемых противоречий новых технологий искусственного интеллекта только усугубило и без того запутанную ситуацию.

Теперь в число взаимодействующих субъектов кроме пациентов, врачей, администраций медицинских учреждений, руководителей системы здравоохранения, вошли компании-производители ИИ, программисты, учёные-исследователи, эксперты и нейросетевые операторы, т.е. все лица, так или иначе задействованные на каком-либо этапе функционирования медицинского ИИ. Наличие столь многочисленного круга участников и отсутствие ясности, не только в отношении того, кто за какую часть конечного результата и в какой мере несет ответственность, но и в отношении того, как на самом деле работает алгоритм нейросети, делает ситуацию абсолютно патовой. В то время как медицинские работники обычно несут конкретную юридическую ответственность за свои действия, что является неотъемлемым элементом их профессиональной деятельности, разработчики ИИ вряд ли сильно заботятся о таких пустяках. Для врача неспособность отчитаться за правомерность своих действий и результаты принятых решений грозит потерей права на работу; в то время как вина разработчика всегда может быть оправдана несовершенством компьютерной техники и языков программирования. Даже если будет установлено, что производитель ИИ на самом деле грубо ошибся, зачастую трудно возложить вину на какого-то определенного человека, поскольку над моделями ИИ работает множество разных специалистов. Кроме того, этические кодексы и системы отчетности, которые используют компьютерные компании, как правило настолько расплывчаты и неконкретны, что их невозможно стандартизировать.

Ещё раз подчеркнём, что подотчетность и ответственность алгоритма медицинского ИИ тесно увязана с его объяснимостью и прозрачностью. Чем не прозрачнее алгоритм, тем труднее найти виновного в нанесении вреда больным. Поэтому основная тяжесть ответственности, скорее всего, будет возложена на врача, который использовал непрозрачный ИИ и, следовательно, не может толком объяснить, как было принято неверное медицинское решение. Это особенно актуально для вспомогательных приложений ИИ, которые создаются для помощи в принятии врачебных решений и поэтому могут рассматриваться как эквивалент консультации с клиническим экспертом. Между тем есть все возможности коренным образом решить проблему подотчетности медицинского ИИ. Во-первых, необходимо четко прописать процедуры определения ролей разработчиков ИИ и клинических пользователей для случаев, когда решения ИИ наносят вред пациентам. Во-вторых, создать регулирующие органы, занимающиеся проблемами медицинского ИИ, которые должны обеспечивать строгое соблюдение нормативной базы и осуществлять привлечение к ответственности конкретных виновников ошибок, включая производителей ИИ [118, 119].

Однако даже если медицинские модели ИИ хорошо проверены и признаны клинически надежными и безопасными, этически обоснованными и полностью соответствующими нормативным требованиям, их путь к внедрению и интеграции в систему здра-

воохранения по-прежнему сопряжен с множеством препятствий. Работники здравоохранения традиционно более консервативны в плане освоения инноваций в сравнении со специалистами других областей народного хозяйства. Кроме того, внедрение ИИ в реальном здравоохранении может также тормозиться из-за: ограниченности структуры и качества данных в существующих электронных медицинских системах, изменчивости отношений между врачами и пациентами, неожиданными трудностями и нестыковками, связанными с клинической интеграцией и функциональной совместимостью различных нейронных сетей.

Начнем с того, что качество электронных медицинских баз данных является ключевым фактором успешного внедрения медицинского ИИ. Однако, как известно, эти данные отвратительно структурированы и чрезвычайно зашумлены, поэтому нейросети практически не могут использовать для обучения львиную долю существующих датасетов. Более того, форматы и качество медицинских данных в разных организациях здравоохранения значительно отличаются. Прежде чем новые технологии искусственного интеллекта смогут быть полноценно внедрены в масштабах государства, накопленные базы данных потребуют значительного и дорогостоящего пересмотра, очистки, дополнительной маркировки и экспертного контроля качества.

Поскольку отдельные медицинские специальности, например, связанные с анализом лучевых изображений, уже претерпели значительные изменения благодаря ИИ, следует ожидать, что нейросети каким-то образом повлияют на отношения врач-пациент. Ориентация технологий искусственного интеллекта на больных может превратить формально патерналистские отношения врач-пациент благодаря повышению прозрачности и доверительности диалога в реальное партнерство при принятии решений. Вместе с тем, нельзя не учитывать этико-деонтологический контекст и последствия ситуаций, связанных со свободным информационным обменом при обнаружении у пациента онкологической или психической патологии. Нужно обязательно обновить клинические рекомендации и протоколы оказания медицинской помощи, чтобы полностью учесть специфику опосредованных искусственным интеллектом взаимоотношений между медицинскими работниками и пациентами [127, 149].

Поскольку поставщики медицинских услуг и врачи работают в соответствии с клиническими рекомендациями и стандартами, внедрение технологий искусственного интеллекта в повседневную практику неизбежно будет иметь практические, технические и клинические последствия как для них, так и для пациентов. Однако пока неясно, насколько модели ИИ будут систематически совместимы с существующими компьютерными системами медицинских учреждений и органов управления здравоохранением и как легко их можно будет интегрировать в текущий производственный процесс без ущерба для такового.



Производителям искусственного интеллекта в сотрудничестве со специалистами по организации здравоохранения будет необходимо разрабатывать стандартизованные регламенты для вновь появляющихся моделей ИИ, чтобы обеспечить их полную совместимость с корпоративными информационными системами медицинских учреждений и быструю интеграцию в гетерогенные электронные базы органов управления страны. В частности, следует разрабатывать новые подходы интеграции нейронных сетей с уже используемыми в медицинской практике технологиями, такими как генетическое секвенирование, электронные карты пациентов и онлайн медицинские консультации [148, 158].

### **Методологические аспекты оценки рисков медицинского искусственного интеллекта**

Описание в предыдущем разделе монографии основных рисков использования медицинского ИИ заставляет нас задуматься о важности структурированного подхода к оценке и управлению рисками, который бы способствовал конкретному решению технических, клинических и этических проблемы применения нейронных сетей в здравоохранении. Риски от медицинского ИИ могут быть охарактеризованы в зависимости от тяжести причинённого вреда, от частоты и вероятности его возникновения. Диапазон рисков начинается от низких рисков, когда пациенту причиняется ограниченный и управляемый вред, и заканчивается высокими рисками, когда человеку нанесён необратимый ущерб. Алгоритм ИИ вполне может отрицательно влиять на производительность врача: например, если модель не может точно очертить границы сердца на объемном изображении, то кардиолог вынужден делать это вручную. Плохо обученный ИИ серьезно навредит больному, если не сможет вовремя диагностировать опасное для его жизни состояние.

Чтобы минимизировать риски ИИ, максимизировав его преимущества в будущем, нужно предвидеть, выявлять, отслеживать и анализировать малейшие потенциальные угрозы в каждом конкретном случае для каждого нового приложения с ИИ. Важным этапом процедуры оценки риска должна стать разработка методологии классификации выявленных рисков по ряду категорий, представляющих различные уровни и виды риска. Для каждого уровня должен быть установлен набор тестов и правил для смягчения и устранения рисков. Более высокие риски требуют большего внимания к тестированию и регулированию, а малые риски – лишь проведения ограниченных мероприятий. Распределение рисков ИИ по степени вероятности, величине и тяжести последствий позволила бы производителям, поставщикам медицинских услуг и регулирующим органам вмешиваться именно в той мере, в какой это необходимо, чтобы обеспечить защиту прав и здоровья пациентов. Главное, чтобы эти классификации не мешали нововведениям в области медицинского ИИ.

В европейских странах уже несколько лет применяются правила по ужесточению требований к клиническим исследованиям, усилению предпродажного контроля и надзора на протяжении всего жизненного цикла устройств, а также накапливается единая база данных эксплуатируемых медицинских девайсов с ИИ. Однако и там не учитываются многие аспекты рисков медицинского искусственного интеллекта, такие как непрерывное обучение нейросетей, выявление алгоритмических ошибок, и тот факт, что ИИ – это высоко адаптивная технология, которая обязательно должна корректироваться по мере поступления новых данных.

Интересно, что одно из первых предложений по оценке рисков в области ИИ появилось в 2018 году, когда в Германии было предложено классифицировать риски общих алгоритмов принятия решений по их критичности, т. е. способности системы причинить вред. Классификация содержала 5 уровней: 1) нулевой потенциальный вред; 2) незначительный потенциальный вред; 3) регулярный потенциальный вред; 4) значительный потенциальный вред; 5) неоправданно высокий вред. В соответствии с этим предложением была предложена адаптируемая в зависимости от уровня риска система тестирования или регулирования, которая предусматривала корректирующие и надзорные механизмы, требования к прозрачности, объяснимости и понятности результатов работы алгоритмов, а также правила распределения ответственности, в том числе, в контексте разработки и эксплуатации нейросетей. В 2021 году Еврокомиссия выпустила предложение по гармонизации правил, регулирующих технологии ИИ во всех странах Евросоюза. Документ содержал оценку рисков ИИ и обязательные требования к платформам ИИ с высоким уровнем риска. Было рекомендовано классифицировать нейросетевые риски ИИ по трём уровням: неприемлемый, высокий, низкий/минимальный [46, 79].

Высшая категория риска соответствует свойствам алгоритма ИИ, которые противоречат ценностям Евросоюза и, поэтому они неприемлемы. Сюда относятся: неявные/неочевидные манипуляции и эксплуатация уязвимостей, наносящие пациенту физический вред или/и психический дискомфорт; навешивание социального ярлыка; несогласованная с пациентом биометрическая идентификация. Промежуточная категория представляет особый интерес и соответствует ИИ высокого риска, который может быть разрешен только в том случае, если будет соответствовать определенным требованиям. Модели ИИ высокого риска должны обязательно содержать компоненты безопасности. К таким моделям относятся: индивидуальные медицинские устройства и более громоздкое оборудование, а также автономные алгоритмы ИИ для эксплуатации критически важной инфраструктуры, для организации частного доступа к медицинским услугам, а также платформы по трудоустройству и управлению персоналом. Помимо этих упомянутых трех категорий рисков (неприемлемых, высоких и низких) в документе обсуждается дополнительная категория нейросетей, для

которых существует жёсткое требование прозрачности, потому что они предназначены: для взаимодействия с людьми, обеспечивать их эмоциональное и биометрическое распознавание. В этом случае люди должны быть персонально уведомлены о том, что им предстоит взаимодействовать с системой искусственного интеллекта.

Между прочим, приятно осознавать, что далеко не каждая модель медицинского ИИ является высоко рискованной. К примеру, в радиологии используются нейросети, обрисовывающие контуры органов и патологических образований (контуры желудочков сердца, границы опухоли легкого). Такие инструменты обработки лучевых изображений на базе ИИ очень важны и полезны для клинической практики. Такие алгоритмы вовсе не обязательно должны быть прозрачными, поскольку врачам достаточно визуально оценить результаты, тут же исправить возможные недочёты, если это вообще потребуется. Здесь риски, естественно, будут минимальны. Поэтому чтобы продолжать продвигать инновации и привлекать инвестиции в медицинский ИИ, требуются чёткие критерии идентификации его уровней риска [91, 113, 167].

Для выявления, прогнозирования и управления рисками в области медицинского ИИ центральное значение имеют адекватные процедуры оценки моделей ИИ. Традиционно оценка ИИ осуществляется в основном путем проверки точности и надежности модели в контролируемых средах. А вот другие аспекты ИИ, такие как клиническая безопасность, эффективность, справедливость, прозрачность, конфиденциальность оценить в лабораторных условиях бывает гораздо сложнее.

Учитывая многочисленные риски ИИ, в том числе и их этические стороны, оценку алгоритмов надо выводить далеко за рамки устоявшихся подходов, которые пока сосредоточены в основном на точности моделей. В полный рост встаёт необходимость в эмпирической оценке производительности алгоритмов машинного обучения в здравоохранении. В таблице 1 приводится перечень элементов производительности, используемых для диагностических алгоритмов в области радиологии. К ним относятся: точность классификации, надежность, применимость, прозрачность, контролируемость, удобство использования и др. (Larson et al., 2021) [99].

Надо честно признать, что представленный выше список не является исчерпывающим, поскольку в нем не учтены некоторые важные риски медицинского ИИ, а именно, алгоритмическая предвзятость и связанная с неравенством несправедливость. Вспомним недавнее исследование, в котором оценивались современные глубокие нейронные сети, основанные на больших данных лучевой диагностики патологии грудной клетки с учетом пола, возраста, расы, а также вида страховки пациента, которая является чётким показателем социально-экономического статуса (Seyyed-Kalantari, et al, 2020). Исследователи пришли к выводу, что модели, обученные на больших наборах данных, естественным образом не обеспечивают равенство возможностей, а вместо этого приводят к потенциальным

**Элементы производительности алгоритмов нейросетевой визуализации.**

Точность	Алгоритм должен точно выполнять все диагностические задачи, для которых он предназначен.
Надёжность	Алгоритм должен оставаться точным в условиях разумно ожидаемых отклонений, встречающихся в клинике, включая допустимые отклонения качества изображений.
Применимость	Точность алгоритма должна поддерживаться для всех методов визуализации и для всех целевых групп пациентов.
Детерминированность	Алгоритм должен давать один и тот же ответ для одного и того же изображения при использовании в разное время и в разных условиях.
Сосредоточенность	Алгоритм должен уметь распознавать важную информацию на изображении, не меняя своей оценки из-за отвлекающих и незначительных деталей.
Опознавание ограничений	Алгоритм не должен выходить за пределы своих возможностей, основываясь на присущих для данной модели ограничениях: клинической применимости; ограничениях, связанных с анатомическими вариациями либо с качеством изображения.
Самоконтроль алгоритма	Алгоритм должен уметь распознавать собственные ошибки и приостанавливать свою работу в момент их обнаружения.
Логическая прозрачность	Пользовательский интерфейс должен позволять оператору четко видеть связь между входящей и исходящей информацией, в том числе: какие данные анализировались, какие варианты рассматривались, почему ИИ отверг те или иные альтернативы, чтобы в конечном итоге иметь убедительные основания для принятия или отклонения выводов алгоритма.



## Элементы производительности алгоритмов нейросетевой визуализации.

Демонстрация уверенности модели	В каждом случае алгоритм должен демонстрировать пользователю степень уверенности в своём заключении. Следует проверять у модели точность оценки её уверенности, равно как и точность работы самой модели.
Мониторинг производительности	Алгоритм должен делиться данными о производительности с пользователями, чтобы обеспечить постоянный мониторинг своих деятельности, быстро выявляя любые существенные отклонения в производительности.
Аудит текущей работы алгоритма	Должны быть предусмотрены независимые средства аудита текущей работы алгоритма, чтобы в критический момент быстро исправить ситуацию. Это может включать периодические проверки качества, аналогичные тем, которые выполняются операторами оборудования для обработки изображений.
Интуитивно понятный интерфейс	Пользовательский интерфейс должен позволять оператору интуитивно понимать, как использовать алгоритм при минимальном обучении и минимальной когнитивной нагрузке.

различиям в медицинском обслуживании, если их использовать без коррекции. В своей работе авторы использовали так называемые истинно положительные показатели *true positive rates (TPR)* в качестве меры справедливости, но в литературе можно встретить и другие критерии: статистическая паритетность *statistical parity*, групповая справедливость *group fairness*, уравненные шансы *equalised odds* и прогнозируемое равенство *predictive equality* [144].

Даже если ИИ будет признан точным, надежным, справедливым и удобным для пользователя, это не обязательно принесет пользу пациенту. Чтобы подтвердить клиническую полезность алгоритма ИИ и позволить ему быть принятым клиническими экспертами, академическими кругами и независимыми общественными организациями, важно оценить его влияние на результаты лечения пациентов. Кроме убежденности в клинической эффективности нейросети, необходимо систематически проводить оценку её экономической эффективности. Учитывая огромные инвестиции

в медицинский ИИ, не следует забывать, что пресловутые эффективность и снижение затрат только предполагаются в будущем. К слову сказать, экономические выкладки с использованием аналитического моделирования принятия решений целесообразно использовать для оценки того, насколько оправданы дополнительные затраты на ИИ, например, в плане его влияния на связанное со здоровьем качество жизни *quality-adjusted life years (QALY)*. Важно отметить, что первоначальные инвестиции и эксплуатационные затраты на сопутствующую инфраструктуру и услуги ИИ должны быть непременно включены в анализ экономической эффективности. Наконец, учитывая, что алгоритмы ИИ продолжают обучаться по мере поступления новых данных, с течением времени важно адаптировать существующие системы проверки, чтобы обеспечить непрерывный мониторинг производительности на протяжении всего жизненного цикла модели ИИ в клинической среде.

Вместо оценки платформы медицинского ИИ в рамках одной процедуры рекомендуется реализовать пошаговый подход, при котором алгоритм проходит несколько этапов оценки с различными целями и возрастающей сложностью. Например, для валидации модели диагностической визуализации достаточно 4 этапа: технико-экономическое обоснование, оценка возможностей, эффективности и устойчивости:

1. Техничко-экономическое обоснование: цель состоит в том, чтобы провести первичную/пилотную оценку алгоритма в лабораторных условиях на тестовом наборе данных. Проводится сравнение с существующими алгоритмами, решающими однотипную клиническую задачу, или с результатами, полученными непосредственно врачами-экспертами. На этом этапе алгоритмы ИИ не обязаны быть полностью надежными, поскольку цель состоит в том, чтобы просто оценить их пригодность. Полученные результаты могут быть опубликованы в научной периодике, даже если алгоритм пока не готов для клинического применения.

2. Оценка возможностей: цель состоит в том, чтобы моделировать реальные ситуации в лабораторных условиях и совершенствовать алгоритм для расширения его возможностей. Этот этап можно считать проверкой *in silico* с использованием компьютерного моделирования или виртуальными клиническими испытаниями. Надежность алгоритма проверяется путем моделирования входных данных и клинических условий, в которых они могут использоваться. Тест на безопасность оценивает способность алгоритма минимизировать риск причинения вреда в непредвиденных ситуациях. Этап проводится с обязательным участием конечных пользователей, особенно врачей и операторов, для оценки их реакции и хода принятия решений в смоделированных условиях.

3. Оценка эффективности: на этом этапе валидация переносится в клиническую среду для оценки реальных показателей и для проведения испытаний на местах – в конкретные медицинские

учреждения. Так можно убедиться, что реальная производительность алгоритма соответствует его производительности в тестовой среде. Все результаты и отзывы этого этапа должны быть использованы для обновления и оптимизации ИИ, который будет повторно протестирован в контролируемой среде, как и на предыдущих этапах, перед следующим раундом клинической оценки на местном уровне. Данный этап оценки выявляет локальные огрехи с контролем качества, поэтому производители ИИ должны работать с местными больницами для решения выявленных проблем с качеством.

4. Оценка устойчивости: на этом этапе производитель должен внедрить механизм, обеспечивающий постоянную оценку и мониторинг производительности с целью постоянного улучшения. Мониторинг и аудит платформы ИИ нужны для автоматического обнаружения ошибок, их исправления, а также для поддержания обратной связи с пользователями. В зависимости от выявляемых проблем алгоритм должен со временем обновляться и улучшаться, в том числе, с использованием дополнительных обучающих данных, а затем повторно тестироваться в контролируемой среде.

При пошаговой оценке успех последующего этапа зависит от успешного завершения предыдущего. Отсутствие в испытательной программе возвратов и повторов сокращает время и снижает совокупные затраты. Кстати заметим, что алгоритмы, которые плохо работают в контролируемой среде, наверняка не смогут хорошо себя проявить и в реальной обстановке. Хотя любые поэтапные испытания и многогранные оценочные валидации требуют сосредоточения усилий и внимания, они очень важны, поскольку учитывают всю сложнейшую палитру проблем, которые встречаются при оказании медицинских услуг под управлением ИИ.

Оценка производительности модели ИИ с использованием наборов данных, аналогичных тем, которые использовались для разработки и обучения модели, называется внутренней проверкой. На заре медицинского ИИ этот способ проверки алгоритмов использовался чаще всего, поскольку его легко реализовать. Однако внутренняя проверка, даже проводимая компаниями с высокой культурой качества, неизбежно окажется предвзятой, а её способность выявлять риски – переоцененной. Поэтому в последние годы производители делают акцент на внешнюю оценку медицинского ИИ. Внешняя проверка подразумевает использование совершенно отдельных внешних наборов данных для оценки инструментов ИИ. Внешние наборы данных должны отражать неоднородность населения и условий использования нейросетей. Такие данные должны собираться из разных медицинских центров и географических местностей, чтобы иметь возможность дать обобщенную оценку работе алгоритма ИИ за пределами контролируемой среды. Благодаря такому подходу можно будет оценить модель в случаях, когда технические параметры сбора данных различаются, например, различное устройство сканеров для визуализации или несов-

падение протоколов ведения больных в разных клиниках. Для внешней оценки и сравнительного анализа платформ ИИ также рекомендуется использовать общие справочные наборы реальных данных, полученные от репрезентативных групп населения. Кроме того, внешняя проверка в идеале должна проводиться с использованием сторонних оценщиков, чтобы обеспечить объективную и исчерпывающую оценку алгоритма ИИ в соответствии с изложенными ранее критериями эффективности, такими как: точность, надежность, справедливость и удобство использования. В число таких сторонних оценщиков могут входить медицинские центры, исследовательские лаборатории или независимые экспертные группы, которые поддерживают стандартные наборы данных. Такая организация тестирования способна обеспечить объективность и самые высокие стандарты качества при оценке и мониторинге решений медицинского ИИ, что приведет к снижению неустраненных рисков и повышению доверия к ИИ в здравоохранении [13, 19, 20, 35].

Для дальнейшего повышения доверия и удобства использования инструментов ИИ необходима подробное документирование и отчетность о ходе проверки. Этот облегчит процесс критической оценки продукта всеми стейкхолдерами и поможет воспроизвести алгоритм и результаты работы ИИ при необходимости. Наиболее известным стандартизированным комплексным руководством по составлению отчетов для используемых в здравоохранении прогностических моделей является *TRIPOD Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis* ([www.tripod-statement.org](http://www.tripod-statement.org)). TRIPOD предоставляет четкие рекомендации о том, что нужно сообщать о прогностической модели, чтобы оценить ее потенциальную погрешность и полезность. В частности, руководство включает в себя контрольный список из 22 пунктов, которые считаются важными для прозрачной отчетности об исследовании модели прогнозирования. Хотя TRIPOD в первую очередь направлен на улучшение отчетности, он также способствует более всестороннему пониманию и анализу моделей прогнозирования, обеспечивая возможность их дальнейшего изучения и использования для руководства оказанием медицинской помощи, тем самым повышая воспроизводимость исследований, доверие к ним. Хотя многие положения руководства по своей сути применимы к исследованиям моделей прогнозирования с использованием методов машинного обучения (ML), его популярность у разработчиков ИИ была очень низкой. Причина этого заключалась в терминологических различиях и кажущейся нерелевантности, поскольку TRIPOD изначально предназначался для прогнозирования на основе регрессии, а не ML. Предполагается, что в скором времени появится расширение TRIPOD-AI, посвященное моделям прогнозирования здоровья, использующим методы ML.

Еще одним примером руководств по отчетности и проверки является *CONSORT Consolidated Standards of Reporting Trials*, имеющее



раздел CONSORT-AI, в котором освещены все специфичные для ИИ элементы и аспекты, в том числе: название, дизайн исследования, участники, результаты и размер выборки, требование, чтобы исследователи предоставляли четкие описания любых вмешательств со стороны ИИ. Также там упомянуто о необходимых для использования нейросети инструкциях и настройках. Исследователи из Стэнфордского университета предложили новый набор стандартов отчетности о решениях ИИ в здравоохранении под названием MINMAR *MINimum Information for Medical AI Reporting*. Стандарты MINMAR описывают минимум информации, необходимой для понимания предполагаемых прогнозов, целевых групп населения, архитектуры модели, процессов оценки и скрытых предубеждений. Рекомендации MINMAR специально разработаны для медицинского ИИ и включают элементы отчетности по трём категориям: общие данные о населении, демографические характеристики пациентов, свойства модели. Такой несложный подход к отчетности по оценке медицинского ИИ, с одной стороны, не исключает прозрачности, тщательности и доверия за счет наличия всей ключевой информации об исследовании в едином документе, а, с другой стороны, будет удобен редакторам публикаций, разработчикам и врачам для быстрого понимания, интерпретации и критической оценки качества дизайна и достоверности результатов исследования [46, 49, 67, 106].

## **2. РАЗРАБОТКА, ОБУЧЕНИЕ И РАЗВЁРТЫВАНИЕ ПЛАТФОРМ МЕДИЦИНСКОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА.**

### **Разработка платформ медицинского ИИ**

Одним из наиболее важных компонентов разработки ИИ для здравоохранения является определение и характеристика проблемы, которую необходимо решить, а затем оценка того, насколько ее можно решить или стоит решать с помощью искусственного интеллекта и ML. Для этого требуется оценка полезности, осуществимости с учетом имеющихся данных, затрат на внедрение, проблем с внедрением в клиническую практику и с эксплуатационным обслуживанием. В этом разделе основное внимание уделяется процессу разработки и валидации модели, а в следующем будут рассмотрены вопросы внедрения, клинического использования и обслуживания.

При рассмотрении полезности медицинской нейронной сети было бы неплохо подумать о том, какие результаты и предпочтения мы надеемся получить. Факторы, влияющие на клиническую полезность прогностической модели, могут включать в себя время выполнения прогноза, наличие смягчающих мер, стоимость и простоту вмешательства, логистику вмешательства и побудительные причины. Хотя оценка модели обычно фокусируется на таких по-

казателях, как: положительная прогностическая ценность, чувствительность или отзыв recall, специфичность и калибровка, ограничения на действие, инициируемое выходными данными модели, когда, скажем, непрерывный мониторинг сердечного ритма может быть ограничен отсутствием достаточного количества холтеровских мониторов. Это нередко встречается в небольших клиниках и поэтому может иметь серьезное влияние на выводы о текущей полезности модели.

Поэтому оценка полезности пары прогнозирование-воздействие от решения ИИ является насущной необходимостью. В ходе процесса оценки важно ответить на несколько ключевых концептуальных вопросов:

- Какими будут последующие вмешательства?
- Кто является целевым пользователем результатов модели?
- Какова последовательность действий при проведении вмешательства?
- Каков риск неудачи и неблагоприятного исхода?
- Каков потенциал вмешательства с учетом имеющихся ресурсов?
- Какая точность необходима и какие ложно положительные или отрицательные результаты менее желательны?
- Каково желаемое изменение результата после вмешательства?

Количественные ответы на эти вопросы могут стимулировать анализ для оптимизации желаемых результатов, корректировки формулировок ожидаемой полезности и фиксации переменных, которые трудно изменить, например, стоимость мероприятия, для получения границ оптимальной полезности.

Для эффективной разработки и проверки приложений ИИ/машинного обучения в здравоохранении необходимо тщательно сформулировать решаемую задачу, принимая во внимание свойства алгоритма, например, положительную прогностическую ценность, и свойства результирующего действия, например, эффективность, а также ограничения на действия, т.е. затраты и возможности, учитывая клиническую ситуацию и психосоциальное окружение. В качестве иллюстративного примера рассмотрим виртуального пожилого пациента, страдающего гипертонией, остеопорозом, диабетом и хронической обструктивной болезнью легких, который обратился за медицинской помощью с жалобами на одышку и сердцебиение. В данном случае лечащему врачу предстоит диагностировать и лечить острое состояние пациента с учетом рисков, связанных с имеющейся хронической патологией. В идеале решение о том, как следует лечить, принимается, исходя из стратификации рисков, при этом решение о выборе методов лечения принимается на основе объективного состояния больного и соответствующих клинических рекомендаций. Проблема в том, что эти рекомендации применяются ко всем группам больных без учета того, была ли данная группа изучена при их разработке. Передовой опыт контроля артериального давления основан на рандомизированных контролируемых исследованиях весьма гомогенного

состава населения, а информация о стратификации рисков развития сердечно-сосудистых заболеваний в зависимости от уровня распространения атеросклероза вообще не поддается обобщению из-за популяционной неоднородности (Yadlowsky, et al., 2018). Более того, недостатки при оказании медицинской помощи могут быть связаны с инерцией мышления, недостаточной осведомленностью врача о передовых методах лечения и игнорированием предпочтений пациентов, что приводит к проблемам в организации ухода и некачественной терапии [167].

Алгоритм ИИ может быть исключительно полезен в лечении таких сложных пациентов, ибо поможет врачу:

- 1) правильно распознать и классифицировать и текущее состояние больного с учетом возраста, расы, наследственности и проч.;
- 2) прогнозировать риски осложнений, например, риск инсульта в течение 1 года с учетом сопутствующей хронической патологии, такой как остеопороз, а не только жалоб больного;
- 3) рекомендовать конкретные методы терапии.

Даже для хорошо изученных клинических ситуаций, таких как инфекционные осложнения после хирургических вмешательств, традиционная практика показала гораздо меньшую эффективность по сравнению с подходом по автоматизированной обработке с помощью ИИ текстового контента электронных медицинских карт, что привело к дальнейшему совершенствованию инструментов ИИ для мониторинга развития постоперационных инфекционных осложнений в режиме реального времени (Ситнова А.В., 2024). Это значительно повышает качество медицинской помощи и экономит время врачей при составлении отчетной документации [32, 166].

Итак, согласно персонализированной оценке риска у нашего виртуального пациента мы опасаемся развития фибрилляции предсердий. Поскольку в отсутствие возможности последующего наблюдения персонализированная оценка наличия невыявленной фибрилляции предсердий не улучшает качества ухода, то проведение холтеровского мониторирования сердечного ритма будет зависеть от наличия нужного оборудования. Известно, что риск развития инсульта и тромбоэмболических осложнений существенно не различается при пароксизмальной, персистирующей и постоянной фибрилляции предсердий. Основное влияние на такой риск оказывают другие факторы: возраст, наличие сопутствующей патологии сердечно-сосудистой системы, сахарного диабета. Для удобства используют оценочную шкалу CHADS2: 1 балл присваивается за наличие хронической сердечной недостаточности, артериальной гипертензии, сахарного диабета, возраст >75 лет; 2 балла – за инсульт в анамнезе. Увеличение оценки на каждый дополнительный балл по CHADS2 сопровождается ежегодным увеличением частоты развития инсульта примерно на 2%: от 1,9% при 0 баллов до 18,2% при 6 баллах (Camm A.J., et al., 2010; Hindricks G., et al., 2021). Если диагноз виртуального пациента был подтвержден и впоследствии шкала риска CHADS2 указала на вы-

сокий риск ишемического инсульта в течение 1 года, полезность лечения антикоагулянтами должна быть определена в свете положительной прогностической ценности шкалы CHADS2 и известной или предполагаемой эффективности антикоагулянтов в предотвращении инсульта, а также с учетом повышенного риска кровотечения при применении антикоагулянтов у больных с артериальной гипертензией [54, 86].

Возвращаясь к вопросу разработки модели медицинского ИИ, нужно отметить, что после того, как установлена её потенциальная полезность, необходимо принять несколько ключевых решений перед началом фактической разработки. Поскольку вариант подхода к моделированию зависит от контекста использования ИИ, на этом этапе, чтобы максимизировать шансы на успех, помимо программистов необходимо привлечение будущих пользователей, ибо клиническая пригодность модели зависит от того, что врач будет с ней делать, т.е. от степени участия и уровня полезности для производственной деятельности.

Ранние автоматизированные экспертные системы занимались накоплением информации и принимали решения посредством явной опоры на знания и опыт экспертов. Эти типы систем использовали устоявшиеся биомедицинские знания, соглашения и отношения. Они управлялись непосредственно экспертами, учитывая нехватку наборов данных для обучения. Современные системы ИИ уже не занимаются переводом предыдущих медицинских знаний в алгоритмы, а пытаются в ходе обучения модели устанавливать смысловые взаимосвязи. Недорогое хранение данных, мощные процессоры и достижения в области машинного обучения сделали возможным глубокую аналитику больших наборов данных, которая помогает выявлять ранее неизвестные науке закономерности. Несмотря на то, что современные нейронные сети во многом превосходят прежние экспертные подходы, однако обучение на реальных данных сопряжено с риском получения потенциально противоречащих здравому смыслу в понимании врачебной аудитории выводов, основанных на слабо интерпретируемых алгоритмах, формированием моделей, которые переобучены и, поэтому лишены способности к обобщению, а также склонны принимать предвзятые решения из-за кодификации бытующих в медицинской практике предубеждений.

Всякий раз, когда алгоритм обучается на основе имеющихся данных или адаптируется к ним, возникает некое противоречие двух противоположных точек зрения на текущее положение дел. С одной стороны, формируется система причинно-следственных связей того, как одни переменные влияют на другие переменные. Таким образом, моделируется основной процесс, который приводит к наблюдаемым результатам. С другой стороны, истинный механизм достижения результатов так и остаётся до конца непонятным, потому что алгоритм полностью сфокусирован на работе математической функции, которая сопоставляет заданный набор входных



переменных с наблюдаемыми результатами. Это означает, что модель просто фиксирует наборы ассоциаций, не гарантируя строгой логической причинности.

Возьмём датасет с набором характеристик от 1 млн человек и попытаемся построить 2 модели для оценки риска сердечного приступа через 1 год. Причинно-следственный подход потребует, чтобы переменные, которые мы собираем, такие как возраст, раса, пол, уровень липидов и артериальное давление, имели некоторую биологическую роль в механизмах, приводящих к сердечным приступам. Как только модель учится на данных, в которых переменные больше всего связаны с результатом, они начинают всё заметнее на него влиять. Изменение этих переменных, например, снижение уровня липидов крови снижает риск сердечного приступа. С позиции формальной логики всё выглядит вполне обоснованно в том смысле, что почему бы уровню риска, определенному однажды в одном из регионов мира, не оставаться таким же и в другом регионе десятилетия спустя. Если мы ставим во главу угла исключительно ассоциации, то тогда в модель можно включить все доступные переменные, например, цвет волос, и даже допустить тот факт, что некоторые важные параметры, такие как упомянутый выше уровень липидов, в датасете отсутствуют. В этом случае алгоритм будет работать как математическая функция, сопоставляющая входные данные с результатами, т.е. будет учиться взвешивать цвет волос как ведущий фактор риска сердечного приступа. Со временем модель придёт к выводу, что седые либо просто светловолосые люди имеют более высокий риск сердечных приступов, хотя окрашивание волос в темный цвет ни на йоту не снижает такой риск. В пользу ассоциативного подхода говорит тот факт, что, если нам нужно оценить финансовый риск при страховании миллионной популяции, когда большая её часть имеет седые или светлые волосы, причинная модель нам вовсе не подойдет. Народная мудрость гласит: «Хорошие предсказания не обязательно должны иметь хорошие объяснения. Сова может быть хорошим охотником, не понимая, почему крыса всегда идет из точки А в точку Б».

В идеале можно пытаться комбинировать подходы, объединяя предварительно накопленные медицинские знания о причинно-следственных связях с изучением ассоциированных данных, чтобы в итоге получить самую точную модель. Такую интеграцию можно проводить творчески, допустим, можно полностью закодировать структуру и устанавливать параметры только на основе данных. Альтернативно можно разрабатывать модели, которые до мелочей отражают то, что известно о заболевании, и уж потом анализировать структуру и оставшиеся неучтёнными параметры.

Обучение модели на предварительных знаниях надежно обеспечивает конструктивную валидность и снижает вероятность нелогичных решений нейросети. Ранее считали использование предварительных знаний или причинно-следственных связей непрактичным для решения любой достаточно сложной проблемы,

отчасти потому, что полный набор причинно-следственных связей редко известен. Однако недавние исследования показали, что даже когда мы установили лишь частичное понимание причинно-следственных связей, мы уже открыли больше возможностей для обобщения, удалив не поддающиеся обобщению ложные связи и, тем самым, снизив риск катастрофически неудачных прогнозов.

Понятие интерпретируемости заслуживает специального обсуждения из-за разночтений в толковании данного термина. Интерпретируемость может означать разные вещи для разработчиков и для пользователей моделей медицинского ИИ. Для программиста интерпретируемость означает способность объяснить, какие переменные или/и их комбинации и каким образом привели к получению результатов. Для врача интерпретируемость может означать одно из двух:

- достаточное понимание того, что происходит, чтобы можно было либо доверять результатам, либо иметь возможность застраховать ответственность за рекомендации ИИ;

- понятность причинно-следственных связей в структуре алгоритма, чтобы иметь возможность мягко корректировать его работу.

Для расчета и оценки 10-летнего риска развития сердечно-сосудистой патологии принято использовать Фрамингемскую формулу *Framingham risk equation*, которая учитывает возраст, пол, уровень систолического артериального давления, общего холестерина и липопротеинов высокой плотности, отношение к курению и наличие сахарного диабета. Фрамингемское уравнение риска удовлетворяет всем трем вышеупомянутым значениям понятия интерпретируемость, хотя она отнюдь не всегда необходима и, очевидно, не достаточна для полезности. При выборе из двух моделей с одинаковой производительностью: 1) модель черного ящика и 2) интерпретируемая модель, большинство пользователей обычно предпочитают интерпретируемую модель. Однако во многих практических ситуациях модели, которые невозможно интерпретировать, могут показывать прекрасные результаты. Совсем не обязательно иметь интерпретируемую модель прогноза дождя, чтобы полагаться на нее при принятии решения о том, брать ли с собой зонтик, если модель хорошо справляется со своими функциями. А ведь прогноз погоды представляет собой сложную задачу, находящуюся далеко за пределами понимания большинства населения. Для неопытных пользователей, если модель добротно работает, степень интерпретируемости совсем не будет влиять на степень доверия ей. Однако искушенные пользователи хотели бы видеть более высокую степень интерпретируемости алгоритма [131].

Во избежание излишних затрат важно понимать, какой уровень интерпретируемости необходим данному конкретному приложению. Рассмотрим в качестве иллюстрации прогноз 24-часовой смертности с помощью глубокой нейронной сети. Для обучения и интерпретации глубоких нейронных сетей может потребоваться довольно много усилий, в то время как линейные модели, которые,

возможно, более интерпретируемы, могут обеспечить удовлетворительную производительность. Если интерпретируемость вдруг считается необходимой, нужно разобраться, почему это так. Однажды был проведен опрос нейросети, предназначенной для выявления причин высокого 24-часового риска смертности больных, на предмет её интерпретации работы собственного алгоритма. В одном из ответов у пациентки с высоким риском суточной смертности в качестве причины установления такого уровня риска было указано наличие метастатического рака молочной железы со злокачественными плевральными выпотами и эмпиемой. Такая интерпретация со стороны нейросети, хотя, по сути, и верна, однако не содержит конкретных рекомендаций, как действовать с учетом полученного прогноза. Модель черного ящика может оказаться достаточной, если результатам доверяют и известно, что рекомендуемое вмешательство влияет на результат. Доверие к результатам можно получить путем тщательного тестирования и перспективной оценки того, насколько часто прогнозы модели являются правильными и калиброванными, а также путем оценки влияния вмешательств на результат. В то же время проспективная оценка может оказаться весьма дорогостоящей. Таким образом, необходимо сделать все возможное для предварительной проверки модели, например, путем анализа изученных ею взаимосвязей.

Среди врачей распространено стремление к коллегиальности и сотрудничеству. В лечебно-диагностическом процессе принимают участие не только врачи, но и множество иных сотрудников, включая средний и младший медицинский персонал, провизоров, социальных работников, различных экспертов и прочих работников системы здравоохранения. Хотя большая часть ведущихся ныне дискуссий об ИИ сосредоточена на том, как инструменты ИИ заменят людей-работников, в обозримом будущем ИИ будет выполнять главным образом дополняющую роль, расширяя возможности медицинского персонала. Поскольку объем информации сегодня нарастает в ускоренными темпами, платформы ИИ вскоре станут незаменимым компонентом производственной деятельности в здравоохранении. Они предоставят специальные знания в предметной области, расширяя возможности врачей и всех участников лечебно-диагностического процесса, делая их работу проще и эффективнее и, в конечном итоге, улучшая уход за пациентами.

В то же время нарастает активная дискуссия об уровне автономии, которую ИИ может или должен иметь в системе здравоохранения. Требуемые характеристики производительности и масштабы использования платформ ИИ существенно различаются в случаях, когда алгоритм работает автономно, или когда врач использует его лишь как ассистента, а окончательное решение принимает сам. Эта коллизия также имеет важные политико-экономические и нормативные аспекты, которые будут обсуждаться нами далее. Вполне вероятно, что прогресс в области медицинского ИИ ускорится, если алгоритмы начнут массово разрабатываться с учетом инте-

ресов всех стейкхолдеров и при их непосредственном участии. Полезность создаваемых на таких принципах экосистем ИИ будет целиком зависеть от их способности расширять возможности человека по принятию решений в области прогнозирования, профилактики, диагностики и лечения заболеваний [5, 7, 112].

## **Обучение моделей ИИ**

Направление ИИ, ориентированное на обучение на основе данных, называется машинное обучение ML. Его можно разделить на 3 основных подхода: контролируемое, неконтролируемое и обучение с подкреплением. Каждый подход может удовлетворить различные потребности в сфере здравоохранения. Обучение с учителем фокусируется на обучении на основе набора помеченных примеров. Каждый пример в лице пациента представлен входящей информацией, включающей: демографические сведения, жизненно важные показатели, результаты лабораторных исследований, а также специальной меткой, например, наличие сахарного диабета. Затем алгоритм проводит сопоставление входных данных с метками, делая обобщения с новыми выводами. Известно множество успешных применений контролируемого обучения в здравоохранении. Rajpurkar с соавт. разработали модель ИИ, которая классифицировала 14 патологических состояний на рентгенограммах грудной клетки не хуже практикующих рентгенологов. На основе обучающих данных алгоритм определял, какие особенности изображения наиболее тесно связаны с различными диагнозами (Rajpurkar, et al., 2018). Такие системы также можно использовать для обучения моделей, предсказывающих будущие события, например, при сопоставлении структурированного и текстового содержания историй болезни с риском смертности, повторной госпитализации и закодированным по выписке с помощью кодов Международной классификации болезней диагнозом. Контролируемый подход к обучению моделей ИИ хорош в условиях с четким результатом и большими объемами размеченных данных. Однако получить размеченные данные не всегда просто. Недвусмысленные метки порой трудно получить по ряду причин: результаты или выводы могут быть неоднозначными при отсутствии консенсуса среди врачей-консультантов, процесс маркировки может оказаться излишне трудоемким и дорогостоящим или проведение разметки может быть просто невозможно. Очень часто в доступе может не быть достаточно большого датасета для уверенного обучения модели. В условиях чрезмерной зашумленности данных предпочтительно обучение со слабым контролем. При необходимости снизить потребность в большом объеме размеченных данных можно использовать альтернативные подходы, в том числе: частично контролируемое обучение или активное обучение [135].

Неконтролируемое обучение направлено на изучение набора немаркированных примеров и их группировку по принадлежности к некоей общности. Кластеризация – одна из распространенных

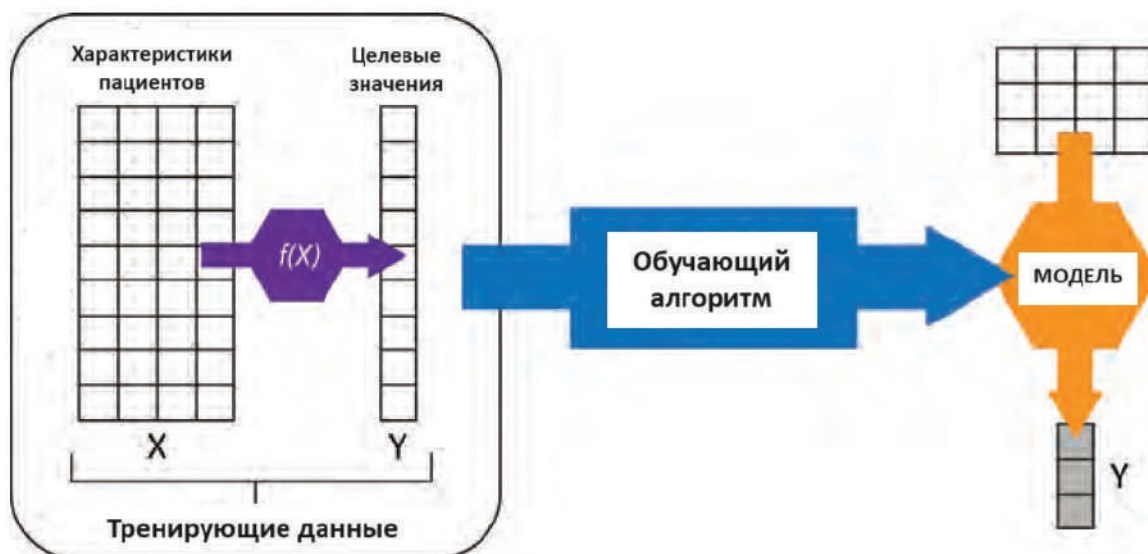


методик обучения без присмотра. Алгоритмы кластеризации в основном используются в исследовательских целях и могут помочь определить структуру и подструктуры данных. Обучение без учителя можно использовать для определения стадии или подтипа гетерогенного заболевания. Здесь трудность заключается не в получении, а в ее оценке группировки, особенно, если данных не очень много. Имея набор данных и алгоритм кластеризации, мы всегда получаем группировку. А проблема состоит в том, является ли наличие групп, т.е. кластеров, действенным свидетельством того, что принадлежность некоего пациента к определенной группе реально влияет на выбор одного из вариантов лечения. Чаще всего способность воспроизводить одни и те же группы в разных датасетах считается признаком того, что группы имеют существенное медицинское значение и с ними следует обращаться особенно внимательным образом. Также если новый объект/факт/запись принадлежит определенной группе, что позволяет приписать ему соответствующий данной группе риск, то это считается признаком того, что такая группировка имеет значение. Например, в выборке из 450 пациентов, страдающих сердечной недостаточностью при сохраненной фракции выброса, приложению с ИИ удалось выделить 3 подгруппы с высоким, средним и низким риском последующей смертности.

Обучение с подкреплением отличается от контролируемого и неконтролируемого обучения, поскольку алгоритм учится посредством взаимодействия со своей средой, а не только с помощью данных наблюдений. Такие алгоритмы в последнее время показывают успехи в игровых сценариях, когда агент, стартуя с начального этапа, совершает действия, влияющие на окружающую среду, и переходя в новое состояние, получает вознаграждение. Эта модель имитирует то, как врачи могут взаимодействовать с окружающей средой, корректируя терапевтическую тактику на основе наблюдаемых эффектов. Обучение с подкреплением наиболее применимо в условиях последовательного принятия решений, когда момент фиксации успеха модели может быть отложен. При обучении с подкреплением большинство приложений ИИ оказываются равно эффективными как в онлайн, так и в офлайн режимах. Обучение с подкреплением является довольно многообещающим, хотя его нынешнее применение не лишено проблем, связанных с путаницей и отсутствием практической целесообразности [146].

Чтобы проиллюстрировать процесс обучения модели, давайте сосредоточимся на контролируемой задаче обучения для стратификации рисков в здравоохранении. Предположим, у нас есть  $n$  пациентов. Каждый пациент представлен  $d$ -мерным вектором признаков, который лежит в некотором пространстве признаков  $X$  (рис.1).

Мы неявно предполагаем, что существует реальная математическая функция генерации данных  $f(x)$ , которая неизвестна и которую мы пытаемся представить с разной степенью точно-



**Рис. 1.** Модель ИИ, представленная в виде карты входных данных ( $X$ ) к выходным данным ( $Y$ ).

сти. У каждого пациента есть метка  $Y$ , обозначающая исход или его состояние, например, наличие или отсутствие диабета. У пациента может быть как одна метка, так и несколько; со временем ситуация может меняться. В самом простом варианте на каждого пациента приходится только одна двоичная метка. Алгоритм должен научиться отображать вектор  $X$  в  $Y$ . Это отображение, составляющее суть модели, выполняется в процессе обучения на базе алгоритма стохастического градиентного спуска.

Степень причинно-следственной связи полученной модели есть степень, в которой она является точным представлением истинного основного процесса, обозначенного  $f(x)$  на Рисунке 1. В зависимости от доступности данных, степени предварительных знаний, используемых для ограничения структуры модели, и используемого алгоритма обучения модели будут поддерживать различную степень причинной интерпретации.

Как только модель обучена на новых данных, представленных векторами признаков, можно оценивать вероятность того или иного результата. Данные, используемые для обучения модели, называются обучающими данными, а новые данные, используемые для оценки эффективности модели, являются тестовыми данными. Обучающие данные также разделяются на подмножества: для обучения и для проверки. Выбор модели, то есть выбор одной конкретной модели из множества возможных с учетом данных обучения, выполняется с использованием тестовых данных.

Познакомьтесь с ключевыми терминами, которые встречаются при изучении темы нейросетевого моделирования:

*Набор обучающих данных:* экземпляр датасета, используемый для обучения модели ИИ.

*Набор данных проверки:* экземпляр датасета, используемый для настройки гиперпараметров модели.

*Набор тестовых данных:* набор данных, который не зависит от набора обучающих данных, но имеет то же распределение, что и набор обучающих данных. Если часть исходного набора данных откладывается и используется в качестве тестового набора, его также называют контрольным набором данных.

*К-кратная перекрестная проверка:* набор данных случайным образом разделяется на  $K$  частей, одна часть предназначается для тестирования, а модель обучается на остальных частях; для оценки модели используется контрольная часть.

*Внешняя перекрестная проверка:* выполняется перекрестная проверка различных настроек параметров модели, наилучший результат фиксируется.

*Внутренняя перекрестная проверка:* выполняется перекрестная проверка данных обучения и модель обучается по наилучшему набору параметров.

*Чувствительность:* доля фактических положительных результатов, которые корректно определены в бинарной классификации. Её также называют истинно положительным показателем *true positive rate (TPR)*, отзывом *recall* или вероятностью обнаружения *probability of detection*.

*Специфичность:* доля фактических негативов, правильно идентифицированных в бинарной классификации. Её еще называют истинно отрицательной ставкой *true negative rate*.

*Точность:* доля прогнозируемых положительных результатов, которые являются истинными положительными. Её также называют положительной прогностической ценностью *positive predictive value*.

*Достоверность:* доля правильно идентифицированных образцов/примеров среди всех исследованных образцов/примеров.

*Кривая рабочей характеристики приемника Receiver operating characteristic (ROC):* кривая, образуемая при построении графика *TPR* в зависимости от частоты ложных срабатываний. Площадь под кривой *ROC* является мерой того, насколько корректно распределены параметры между двумя различными группами.

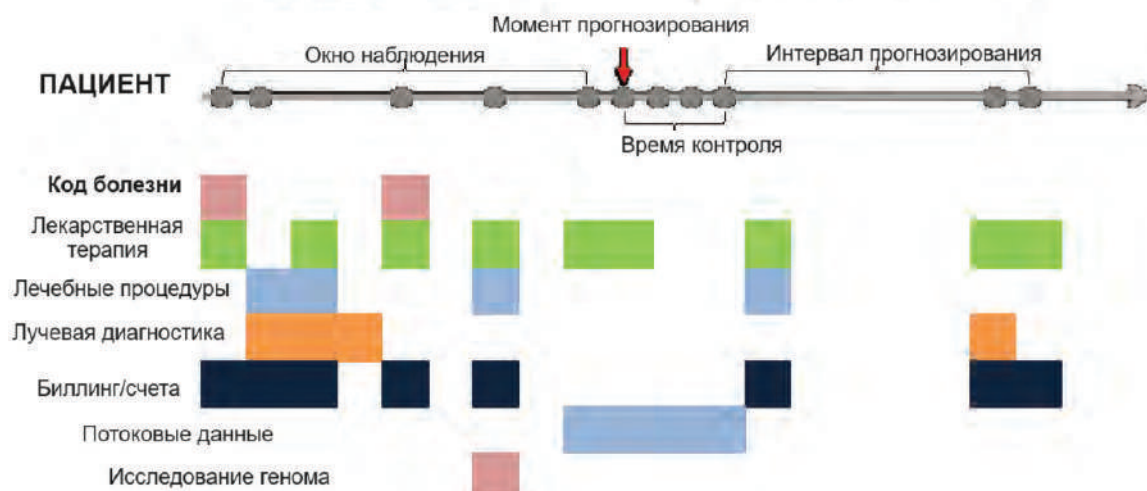
*Кривая точности отзыва Precision-recall (PR):* кривая, образуемая при построении графика зависимости точности от отзыва, чтобы показать компромисс между ними при различных настройках параметров. Площадь под кривой *PR* является наилучшим показателем для сильно несбалансированных задач классификации.

Неправдивая и некачественная информация для обучения приводит к созданию плохих моделей, что напоминает старую истину: мусор на входе – мусор на выходе. Не понятно откуда пошла мода превозносить ИИ как нечто волшебное, способное обучаться независимо от входных данных. На практике **выбор данных всегда важнее выбора конкретного алгоритма!!!** При выборе данных для любого проекта по моделированию интересующий результат, например, смертность в стационаре и процесс его получения, скажем, определяемый с помощью анализа диаграмм сводной записи о выписке, должны быть описаны воспроизводимым образом. Если проблема связана с вре-

менными рядами, время, в которое результат будет наблюдаться и записываться, по сравнению со временем, когда его необходимо спрогнозировать, должно быть определено заранее (рис. 2). Также наряду с данными окна наблюдения, использованными для изучения модели, должно быть учтено время выполнения прогноза.

Отдельные события на временной шкале обозначены серыми кружками. Цветные части временной шкалы под серой линией показывают различные типы данных, которые могут быть собраны при разных встречах, а также тот факт, что не все собираются одновременно. Практически ни один источник данных не обеспечивает непрерывное слежение за состоянием здоровья пациента, за исключением потоковых данных с мониторов отделения интенсивной терапии, используемых в короткие промежутки времени. Носимые устройства также обеспечивают непрерывное поступление данных, но их использование в отечественном здравоохранении пока широко не распространено. Красная стрелка показывает выбранную точку на временной шкале, где предпринимается попытка прогнозирования. Для обучения модели для этого прогноза доступны только данные до этого события. Каждое предсказание дает возможность предпринять определенные действия до того, как предсказанное событие произойдет. Интервал времени между данными прогнозирования и самым быстрым возникновением прогнозируемого события указывает время, доступное для выполнения необходимых действий по смягчению последствий.

Важно иметь подробное описание процесса сбора данных, критерии выбора обучающих данных, а также описание и распространенность атрибутов, которые могут повлиять на то, как модель будет работать с новым датасетом. При построении прогностической модели может возникнуть систематическая ошибка отбора, когда субъекты в обучающих данных являются нерепрезентативными для целевой совокупности. Также ошибки в измерении воз-



**Рис. 2.** График лечения пациента, привязанный к процессу сбора информации



действия или заболеваемости могут стать источником систематической ошибки измерения, особенно при использовании данных электронных историй болезни в качестве источника информации. Как ошибки выбора, так и ошибки измерения могут повлиять на точность прогнозной модели, а также на результаты обобщения полученных данных.

Степень, в которой обучающие данные влияют на корректность обобщений алгоритма ИИ, зависит от метода, используемого для моделирования, и систематических ошибок, присутствующих во вводимых данных. Поэтому алгоритмы могут наследовать многие родимые пятна сложившейся в учреждении практики оказания медицинских услуг. Большинство моделей, обученных с учителем, в новой среде будут вести себя подобно тому, как они привыкли действовать в среде разработки. Левые и правые цензурные настройки или их отсутствие также могут влиять на качество обобщений. Нужно помнить, что процессы генерации и сбора данных со временем претерпевают изменения. Это явление, которое носит название *нестационарность данных*, существенно вредит производительности алгоритма. Модели, обученные с использованием устаревших данных, рискуют попасть под санкции за предвзятость и несправедливость.

Решения, принятые во время создания и ввода наборов данных, будут отражаться на результатах работы будущей модели. Помимо знания окончательных характеристик, представляющих данные пациента, любые этапы предварительной обработки должны быть четко задокументированы и доступны пользователям. Такие важные шаги в обработке данных, как приёмы обращения с пропущенными значениями или нерегулярными выборками, часто упускаются из виду или замалчиваются. Выбор, сделанный при подготовке данных и преобразовании их в аналитическое представление, может в значительной степени способствовать инкорпорируемой в алгоритм ИИ предвзятости.

Пользователи моделей ИИ порой необоснованно возлагают ответственность за предвзятости на алгоритм, а не на загруженные в модель данные. Известны случаи, когда алгоритмы транслировали предвзятости, присущие используемым для их обучения данным. Например, попытка использовать данные Фрамингемского исследования сердца *Framingham Heart Study* для прогнозирования риска сердечно-сосудистых катастроф среди меньшинств привела к предвзятым оценкам риска. Трудности с предвидением неявной дискриминации в сфере здравоохранения приводят к попаданию такого типа предвзятости в обучающие датасеты платформ ИИ (Shah, et al., 2018). Подобные предубеждения чреваты тиражированием неверных решений: если врачи часто отказывают в помощи пациентам с угрожающими жизни состояниями, например, крайне недоношенным младенцам или пострадавшим с отрывной черепно-мозговой травмой и выраженной кровопотерей, платформы машинного обучения могут прийти к выводу, что такой подход

является единственно правильным, а исходы всегда фатальны. Проблема здесь очевидна – наличие прямой причинно-следственной связи между навязанным модели предубеждением и ошибочным выбором алгоритма в отношении лечебной тактики [145].

Часть набора обучающих данных обычно отводится для настройки гиперпараметров, т.е. весов, присвоенных различным переменным и их комбинациям. Эта часть обучающих данных называется набором данных проверки гиперпараметров или просто набором проверочных данных, который нужен для подтверждения правильности выбора значений параметров модели. Необходимо отметить, что здесь речь идет о валидации именно таких данных. Их не нужно путать с результатами клинической валидации или внешней валидности. Учитывая, что модель разрабатывалась на основе данных обучения/проверки, необходимо оценить ее эффективность при классификации или прогнозировании путем тестирования, на датасете, который ранее не использовался бы для выбора параметров модели или гиперпараметров. Тестовый датасет должен быть приближен к данным, с которыми модель будет иметь дело в процессе эксплуатации. Выбор показателей, используемых для оценки производительности модели, определяется конечной целью моделирования, а также типом проводимого обучения. Следует фокусироваться на показателях для контролируемых бинарных классификаторов типа инструментов стратификации риска пациентов. Оценку показателей лучше производить посредством перекрестной проверки, при которой весь набор данных случайным образом разбивается на несколько частей, причем одна часть используется в качестве тестового набора, а остальные части для обучения модели.

Отклик/полнота *recall* и точность *precision* – 2 наиболее популярных у разработчиков показателя производительности нейросети, имеющих различную значимость в зависимости от варианта использования. *Recall* демонстрирует способность алгоритма обнаруживать данный класс вообще, а *precision* – способность отличать данный класс от других классов. Чувствительность *sensitivity* же количественно определяет способность классификатора идентифицировать истинно положительные случаи. Обычно высокочувствительный классификатор может надежно исключить заболевание, если его результат отрицательный. Точность количественно определяет способность классификатора правильно идентифицировать истинно положительный случай, т.е. она оценивает количество раз, когда классификатор ошибочно принимает событие с нулевым значением за реальный случай. Специфичность *specificity* определяет количество фактических негативов, которые правильно идентифицированы как таковые. Существует компромисс между показателями полноты, точности и специфичности, который необходимо решить на основе интересующего нас клинического вопроса. В ситуациях, когда мы не можем позволить себе пропустить случай, желательна высокая чувствительность. Часто высокочувствительный

классификатор сопровождается высокоспецифичным тестом для выявления ложноположительных результатов среди тех, которые отмечены чувствительным классификатором. Компромисс между специфичностью и чувствительностью можно визуально оценить по кривой рабочей характеристики приемника *receiver operating characteristic (ROC)*. Площадь под кривой *the area under the ROC (AUROC)* служит для обобщения информации в кривых ROC. Собирая о результатах контрольного набора тестов, рекомендуется не ограничиваться индексом *AUROC*, а анализировать всю кривую ROC, учитывая чувствительность, специфичность, положительную прогностическую ценность и отрицательную прогностическую ценность в различных точках кривой, которые в купе объективизируют разумность принимаемых платформой ИИ решений.

Несмотря на популярность кривой ROC и *AUROC* и для оценки производительности классификатора и то, что ограничения ROC-кривых хорошо известны, немаловажно упомянуть о следующем:

1. Полезность применения ROC и *AUROC* может сильно варьировать, т.к. возможно, что в зависимости от формы кривой ROC классификаторы с более низкими значениями *AUROC* могут оказаться более полезными, чем те, что имели более высокие значения;

2. В сильно несбалансированных наборах данных, где отрицательные и положительные метки распределены неравномерно, кривая точного отклика *precision-recall curve (PR)* объективно лучше подходит для сравнения классификаторов. Следовательно, чтобы обеспечить содержательность сравнительного анализа, разработчики программного кода должны обнародовать показатели площадей под кривыми и *AUROC*, и *PR*, а также фактические данные с распределением ошибок вокруг средней производительности классификатора.

Для принятия решений в клинике дополнительно необходимы такие показатели, как: калибровка *calibration*, чистая реклассификация *net reclassification* и оценка полезности *utility assessment*. В то время как кривые ROC предоставляют информацию о способности классификатора отличать истинный случай 1 от нулевого события 0, метрики калибровки количественно определяют, насколько хорошо предсказаны вероятности того, что истинный случай на самом деле является 1, согласуются с наблюдаемыми пропорциями случаев и нулевых событий. Для хорошо откалиброванного классификатора в 90 из 100 образцов с прогнозируемой вероятностью 0,9 должны быть правильно выявлены все истинные случаи (Lorent, et al., 2019) [109].

При оценке потенциала моделей машинного обучения также важно опираться на сопутствующие базовые модели типа штрафной регрессии *penalized regression*, тестируемые теми же наборами данных, как и более сложные модели: глубокое обучение *deep learning* или случайные леса *random forests*. Учитывая неочевидную связь между положительной прогностической ценностью *positive predictive value*, полнотой *recall*, специфичностью *specificity* платформы и ее полезностью *utility*, наличие этих параллельных моделей предоставляет возможность применения ещё одной

шкалы для оценки алгоритма с точки зрения стоимости реализации, интерпретируемости и относительной производительности.

Помимо вопросов количественной оценки полезности платформы медицинского ИИ, существует ряд методологических проблем, связанных с необходимостью постоянной оценки производительности нейросети по мере изменения датасета, лежащем в основе её работы. Допустим, модель прогнозирования 24-часовой смертности нужно переобучать каждую неделю или даже каждый день по мере поступления новых данных. непонятно, какие ключевые данные и показатели производительности нам следует использовать, чтобы управлять подобными постоянно развивающимися платформами. Также неясно, каков должен быть график переподготовки, и какая информация должна лежать в основе нашей стратегии. Подробности подходов к надзору и реализации моделей будут подробно рассмотрены ниже. Также мы обязательно коснёмся юридических аспектов тех проблем, которые возникают, если модель после переобучения вдруг начинает вести себя необычным образом.

На целостность и качество данных в здравоохранении влияет целый ряд факторов. Например, программное обеспечение для поиска, предварительной обработки и очистки данных часто теряется или не обслуживается, что делает невозможным воссоздание того же набора данных. Кроме того, исходные из каких-либо источников могут быть просто удалены или изменены. Ведь с течением времени происходят большие и малые изменения в институциональных хранилищах данных или процедурах управления. Наконец, разрозненность знаний и доступ к источникам информации создают зависимость от отдельных людей или целых коллективов. Когда сбор и происхождение данных, на которых обучается модель, представляют собой черный ящик, исследователи должны компенсировать эту зависимость от доверенных лиц, что неоптимально и ненадёжно в долгосрочной перспективе. Разработка алгоритмов на основе плохих данных только усиливает негативные реакции общественности на ИИ. Помните, что информация об этнической принадлежности и имущественном статусе человека не регистрируется и отсутствует в базах данных медицинских центров. Учитывая вышесказанное о низком качестве исходных данных, вопрос о несправедливости моделей прогнозирования в отношении этнических и социально незащищенных групп населения можно считать озвученным. Медицинскому сообществу необходимо вплотную озаботиться улучшения качества данных, которые собираются системой здравоохранения. Чтобы эффективно использовать ИИ, важно следовать передовой практике работы с информацией как при создании и хранении ретроспективных датасетов, так и при перспективном сборе данных, поскольку качество и объем данных прямо влияют на успешную разработку и эксплуатацию моделей.

Итак, мы выяснили, что для успешной разработки платформ ИИ необходимы высококачественные данные. Однако процедуры



оценки качества имеющихся данных и методология создания высококачественного датасета не стандартизированы или вообще отсутствуют. Методы оценки достоверности и воспроизводимости данных часто носят закрытый характер. В идеале данные должны подвергаться перекрестной проверке, чтобы убедиться в их надежности. В проверках данных как по результатам, так и по объясняющим переменным, должны участвовать несколько независимых экспертов, работающих в данной области. Привлечение нескольких компетентных специалистов по ручному абстрагированию и аннотированию данных гарантировано обеспечит правильную оценку присущих данным неоднозначности и изменчивости. Например, при подготовке датасета для обучения платформы ИИ для анализа инфекционных осложнений после хирургических вмешательств у разработчиков не возникло никаких сомнений в отношении наличия инфекционного агента, а вот единодушия в оценке причин, тяжести и прочих характеристик выявленных осложнений достичь так и не удалось (Ситнова А.В., 2024). Недостаточное отражение происхождения и семантики каждого конкретного случая в наборах данных по крайней мере контрпродуктивно. В худшем варианте это может быть крайне опасно и вредно, поскольку наборы данных могут иметь явные предвзятости или расплывчатые предположения, что приводит к формированию алгоритма, который будет уверенно выдавать контекстуально зависимые или откровенно ложные результаты. В конечном счете, чтобы предсказаниям или рекомендациям медицинских платформ ИИ можно было доверять, семантика и происхождение исходной информации должны быть полностью прозрачны, недвусмысленны и доступны для проверки в клинических условиях [32].

Часто упускаемая проблема, связанная с данными, заключается в том, что данные, используемые для обучения модели, должны быть такими, чтобы они были действительно доступны в реальной среде, где будет использоваться ИИ, обученный на этих данных. Например, ИИ, анализирующий сигналы электрокардиограммы, должен иметь возможность доступа к ним в пункте регистрации. Часто записи с холтеровских мониторов ждут клинической интерпретации в течение нескольких часов, если не дней, из-за сложности обработки, тогда как патологические изменения на электрокардиограмме могут быть быстро идентифицированы нейросетью для принятия неотложных мер и спасения жизни больного. В связи с этим в число разработчиков ИИ должны входить инженеры, способные налаживать непосредственную передачу сигналов ЭКГ-мониторов в конструируемую модель.

В дополнение к вышесказанному, ещё одним немаловажным требованием является возможность получения составных элементов обучающих данных без особых усилий. Модели, обученные на наборах данных, которые абстрагировались вручную, например, не могут быть развернуты без дополнительных инвестиций со стороны ранее приобретшего их обладателя, которому вдруг

понадобились данные о пациентах, в интересах которых он начал эксплуатировать данную модель. В принципе эту проблему можно решать с помощью методов компьютерного фенотипирования. Однако этот метод имеет ограничение, связанное с несовместимостью корпоративных информационных систем медицинских организаций, что приводит к несостыковкам в формах отчетности и различиям в организации лечебно-диагностического процесса. С появлением стандартов совместимости типа *Fast Healthcare Interoperability Resource*, масштабы этой проблемы со временем, вероятно, уменьшатся. Когда вычислительные фенотипы начнут служить основой для последующей аналитики, будет нужно, чтобы сами они были четко определены, адекватно отражали целевую область и хорошо управлялись.

Для минимизации проблем с качеством и увеличения ценности данных они должны соответствовать принципам FAIR: находимость / *findability*, доступность / *accessibility*, совместимость / *interoperability* и возможность повторного использования / *reusability*. Исследователи в области молекулярной биологии и биоинформатики выдвинули эти принципы, хотя, по признанию многих специалистов в области цифрового здравоохранения, применять комплект FAIR будет весьма непросто. Досадным препятствием для эффективного продвижения и набора популярности медицинского ИИ сегодня является разрозненность представленных в отечественном здравоохранении типов данных. Электронные истории болезни, носимые устройства и гаджеты значительно увеличили объем, разнообразие и скорость накопления клинических данных. Перспектива скорого появления и использования врачами геномной информации только усугубляет проблемы, связанные с установлением происхождения, определением достоверности и чистоты данных, а также с обеспечением своевременной доступности к ним заинтересованных пользователей.

Мы рекомендуем никогда не забывать о временной шкале формирования медицинской документации пациента, которая была представлена на рисунке 2. Там должно быть зафиксировано точное время появления различных типов данных и их происхождение. Само собой разумеется, что любая прогностическая или классификационная модель, действующая в какой-то конкретный момент временной шкалы, может рассчитывать только на обработку тех данных, которые успели накопиться до этого момента. Именно такое реальное представление о доступности данных имеет решающее значение при подготовке платформ ИИ, поскольку использование чистых данных дает слишком оптимистичный взгляд на их эффективность и создает иллюзорное представление об их потенциальной ценности. Интересно, что стратегия использования синтетических данных, если они создаются, чтобы быть виртуальными копиями-заменителями реальных данных с учетом отсутствия либо задержек с их получением, может быть весьма полезным подспорьем разработчикам для создания реалистичных сред обучения

и тестирования для новых алгоритмов ИИ (Carnegie Mellon University, 2018; Schuler, 2018) [57, 148].

Критически важно незамедлительно начинать массовую подготовку студентов медицинских университетов в области науки о данных / data science и искусственного интеллекта. Прогресс в области цифровой медицины прямо зависит от накопления критической массы специалистов с медицинским образованием, профессионально разбирающихся в этих областях. Ахиллесовой пятой крупных чиновников от медицины из высоких кабинетов, принимающих судьбоносные решения и курирующих инновационные проекты ИИ, является их некомпетентность, порождающая упорное желание получить максимальный результат за минимальный срок и громко раструбить об ошеломляющих успехах. Специалисты по ИИ ощущают невероятное административное давление сверху, которое обычно приводит к двум крайностям. С одной стороны, стремительное внедрение сырых решений искусственного интеллекта может привести к вводу в эксплуатацию алгоритмов без адекватного понимания врачами их возможностей влиять на лечебный процесс и ограничений производительности, что грозит напрасной тратой ресурсов учреждения и нанесением пациентам непоправимого ущерба. С другой стороны, наделение платформ ИИ сверхчеловеческими способностями по мгновенной обработке огромных информационных массивов, обуславливающее иллюзорное превосходство нейросетей над авторитетом врача, также является опаснейшим заблуждением, которое может привести к тому, что передовые решения будут с ходу отвергнуты врачебными коллективами. Производителям программного обеспечения нередко свойственно переоценивать пользу своих разработок в сфере искусственного интеллекта. Они абсолютно не склонны уделять много внимания снижению рисков причинения вреда больным и экономичности работы платформ ИИ. Чрезмерные финансовые, временные и психоэмоциональные затраты пользователей их мало интересуют. А вот собственные расходы они с удовольствием снижают, применяя тестовые шкалы производительности моделей, никак не коррелирующие с запросами пользователей, либо методы оценки, не ориентированные на простые и очевидные сравнения исходных данных (Christodoulou, et al., 2019) [59].

В настоящее время большая часть существующих дискуссий сосредоточена на оценке моделей ИИ с технической точки зрения. Парадоксально, но чисто экономические расчёты прибыли от интеграции ИИ в производственные процессы учреждений здравоохранения пока сильно не афишируется, оставаясь как бы за кадром. После того, как потенциальная полезность установлена, разработчикам и пользователям модели необходимо тесно сотрудничать в период её обучения, поскольку варианты подходов к моделированию зависят от контекста будущего использования нейросети. Например, внешняя валидность во многом зависит от того, что пользователи собираются делать с моделью, т.е. степени

пригодности модели и характера инициируемых ею действий медицинского персонала.

Очередной раз предупреждаем: лукавые данные приводят к созданию лукавых моделей, поэтому датасет, который выбирается для обучения, имеет несравнимо большее значение, чем конкретный математический алгоритм. Программистам стоит основное внимание уделять обучающим данным, а не думать только о технической стороне разработки. Даже при технической оценке необходимо выйти за рамки кривых ROC и изучить множество аспектов эффективности, о которых уже было рассказано ранее. Мы говорили и будем говорить, что для принятия решений на уровне клинической апробации алгоритма необходимы дополнительные показатели, такие как: калибровка / *calibration*, улучшение чистой реклассификации *NRI / net reclassification improvement index* и оценка полезности / *utility assessment*. Учитывая неочевидную взаимосвязь между положительной прогностической ценностью, полнотой, специфичностью модели и ее полезностью, важно проанализировать ряд сопутствующих базовых показателей, таких как: регрессия со штрафом, применяемая к аналогичным данным, которые предоставляются и в более сложные модели с глубоким обучением.

Вопросы интерпретируемости алгоритмов по праву заслуживают настойчивого упоминания на страницах нашей книги из-за нестихающих дебатов вокруг этой темы. Для создателя модели ИИ интерпретируемость часто означает способность объяснить, какие переменные, их комбинации и каким образом ведут к получению тех или иных результатов. Для клинициста же интерпретируемость может означать одно из двух:

- 1) ясное понимание принципов работы алгоритма для того, чтобы можно было доверять результатам ИИ и иметь возможность устанавливать ответственных за ошибочные рекомендации;

- 2) устойчивость причинно-следственных связей в структуре алгоритма, чтобы можно было своевременно предпринимать необходимые корректирующие действия при отклонениях и сбоях.

Важно заранее знать, какая интерпретируемость требуется в каждом конкретном приложении, чтобы избежать лишних проблем при её эксплуатации. Модели черного ящика может быть вполне достаточно, если врачи ей доверяют. Такое доверие можно обрести демонстрацией проспективной оценки того, что рекомендации модели верны и четко откалиброваны.

Низкое качество данных отрицательно влияет на уход за пациентами и его результаты. Платформы ИИ могли бы быть значительно улучшены за счёт выполнения следующих требований: использование многоцентровых датасетов, включение в обработку изменяющихся во времени данных, оценка недостающих данных, а также цензура информации и разработка показателей клинической полезности. В качестве ориентира в решении проблемы качества данных и максимизации их ценности, они должны соответствовать



принципам FAIR, о которых ранее уже упоминалось. К сожалению, часто упускается из виду, когда и где становятся доступными определенные данные и обеспечен ли платформе механизм доступа к датасету. Самым актуальным является создание многоуровневой системы обучения всех заинтересованных лиц, начиная от разработчиков платформ медицинского ИИ и заканчивая простыми пользователями. Наступление нейросетей по всем фронтам сопряжено с широким спектром правовых и этических проблем, которые все еще ждут своего решения. Когда врач предпринимает некие действия на основе рекомендаций ИИ, совершенно не понятно, на кого следует возлагать вину в случае неудачи. Надо заметить, что большинство подобного рода проблем никогда в истории развития компьютерной техники не были решены продумано и окончательно. Новая волна мощнейших алгоритмов, основанных на искусственном интеллекте (вспоминаем о пресловутом чёрном ящике), только усугубляет сложившуюся на сегодня патовую ситуацию.

Разработчикам моделей нужно тщательнейшим образом подходить к выбору обучающих датасетов, а руководители учреждений здравоохранения должны научиться выходить за рамки клише о полезности ИИ, обращая фокус внимания на технические характеристики платформ. Медиа ресурсам необходимо взвешенно и всесторонне освещать как огромный потенциал, так и риски, связанные с использованием медицинского ИИ. Важно продвигать сбалансированный подход к внедрению нейросетевых технологий, который будет способствовать расширению роли ИИ как помощника, а не как способа замены врача. Это позволит медицинскому сообществу лучше понимать, когда и как использовать интерпретируемые модели вместо моделей черного ящика, а также позволит избегать многих этических коллизий, которые неизбежно возникнут по мере широкого распространения медицинского ИИ [61, 105].

### **Развертывание платформ медицинского ИИ в клинике**

В настоящее время нарастающая популярность медицинского искусственного интеллекта во всем мире даёт нам повод глубоко задуматься над тем, почему до сих пор мы наблюдаем откровенно мало примеров развертывания и использования нейросетевых платформ в системе отечественного здравоохранения, а также не видим достаточно большого количества примеров, убедительно доказывающих реальное улучшение качества и доступности медицинской помощи, достигнутое после развертывания платформ ИИ. В частности, существование обширной литературы по разработке и проверке моделей прогнозирования рисков машинного обучения резко контрастирует со скудными данными, описывающими успешное клиническое внедрение этих моделей в учреждениях здравоохранения. Аналогичная картина характерна не только для России, но и для экономически более развитых стран западного мира (He et al., 2019) [85].

Бьющее в глаза несоответствие между усилиями по разработке и успешным использованием ИИ отражает препятствия на пути более масштабного развертывания платформ медицинского ИИ и инструментов поддержки принятия решений на его основе. Хотя некоторые препятствия носят чисто технический характер, большая их часть связана со сложностью интеграции приложений ИИ в существующие корпоративные информационные системы учреждений здравоохранения, полным непониманием потребностей и ожиданий медицинского персонала на местах, неопределенностью целевых установок самих платформ ИИ, отсутствием транспарентности производственных процессов медицинских организаций и вполне обоснованными опасениями врачей в отношении юридической ответственности. Теоретически указанные препятствия могли бы быть уравновешены потенциальными выгодами, поскольку только поверхностный обзор судебных исков по поводу врачебных ошибок показывает, что более половины из них можно было бы предотвратить с помощью хорошо продуманной системы поддержки принятия клинических решений *CDS clinical decision support*. Это могут быть напоминания в форме звуковых и письменных оповещений о лекарственной несовместимости, об отклонениях в анализах биологических жидкостей и т.п. Хотя внедрение инструментов ИИ в отечественное здравоохранение пока нельзя считать массовым явлением, вызывает беспокойство, что более 50% начинаний в области нейросетевых технологий терпят коммерческую неудачу.

Если оставить в стороне проблемы поддержки принятия врачебных решений, у ИИ есть прекрасная возможность найти применение в областях, выходящих за рамки лечебно-диагностического процесса, например, в медицинском менеджменте на уровне групп населения или популяции, где данные и стандарты работы могут быть более объективно определены. Смена приоритетных направлений атаки нейросетевых технологий, скорее всего, будет сопровождаться совсем другими трудностями, связанными с преобразованием приложений ИИ в действенные инструменты, нацеленные на повышение качества и эффективности работы системы здравоохранения в целом. Инструментам ИИ свойственно порождать проблемы, связанные с незрелостью технологии. Уже на самых ранних стадиях внедрения нейросетей в здравоохранение использование инструментов ИИ возбудило опасения врачей и руководителей здравоохранения в отношении: прозрачности моделей, правдивости лежащих в их основе данных, применимости приложений ИИ в открытии новых причинно-следственных связей, этики и деонтологии их применения. На данном этапе цикла разработки методы оценки требований и обслуживания инструментов ИИ, а также лежащих в их основе потребностей в данных до сих пор остаются на рудиментарной стадии.

Поддержка принятия решений обычно подразумевает предоставление четких рекомендаций, касающихся диагностики заболеваний

и/или прогнозов предложенной терапии, направленных на удовлетворение признанной потребности в медицинской помощи, а также на содействие процессам, обеспечивающим безопасность лечебных мероприятий и оперативных вмешательств. В настоящее время большинство врачей регулярно сталкиваются с электронными историями болезни / *EHR electronic health record*. Не входящие в EHR приложения для оказания медицинской помощи называются программным обеспечением как медицинское устройство / *SaMD software as a medical device*. SaMD с ИИ, как любая другая нейросетевая платформа, нуждается в тщательном многократном тестировании, повторных калибровках и проверках в ходе развертывания, а также периодически и после такового. Все приложения медицинского ИИ на местах должны соответствовать передовым практикам в отношении формы, функций и размещения в рабочем процессе систем поддержки принятия врачебных решений *clinical decision support CDS*, а также включать лучшие образчики взаимодействия человека с компьютером и проектировочных решений влияния человеческого фактора. Это должно происходить в среде активно развивающегося электронного документооборота, где присутствуют перспективы использования технологий ИИ.

Личное общение с пациентами в некотором смысле является лишь верхушкой айсберга всего комплекса организации медицинской помощи, для оказания которой требуется множество специалистов и процессов. Весь этот конгломерат постоянно генерирует и потребляет огромные объемы информации. Результаты диагностики и лечения пациентов, их интерпретации и рекомендации по уходу должны своевременно и достоверно сообщаться в рамках установленных ограничений всем заинтересованным сторонам. Безусловно, нейросетям предстоит играть ведущую роль в распознавании изображений в радиологии, офтальмологии, дерматологии и патанатомии, а также обработке сигналов от электрокардиографов, электроэнцефалографов и при проведении аудиометрии. Помимо интерпретации изображений, ИИ будет использоваться для интеграции и сопоставления результатов тестов и анализов с прочими клиническими данными для повышения качества медицинской помощи.

Административно-хозяйственные информационные системы медицинских учреждений по мере увеличения количества и сложности инструментов ИИ будут значительно совершенствоваться. Инвесторы и новаторы в сфере здравоохранения делают ставку на завоевание искусственным интеллектом ведущих позиций в ведении бэк-офисной деятельности, поскольку она несколько легче формализуема, чем лечебно-профилактическая работа, и обычно представляет меньший риск для жизни и здоровья пациентов. Кроме того, необходимые для обучения моделей данные из административно-хозяйственной сферы гораздо более доступны и достоверны, чем чисто клинические датасеты. В больницах административные задачи включают: управление документооборотом, аптекой, ресурс-

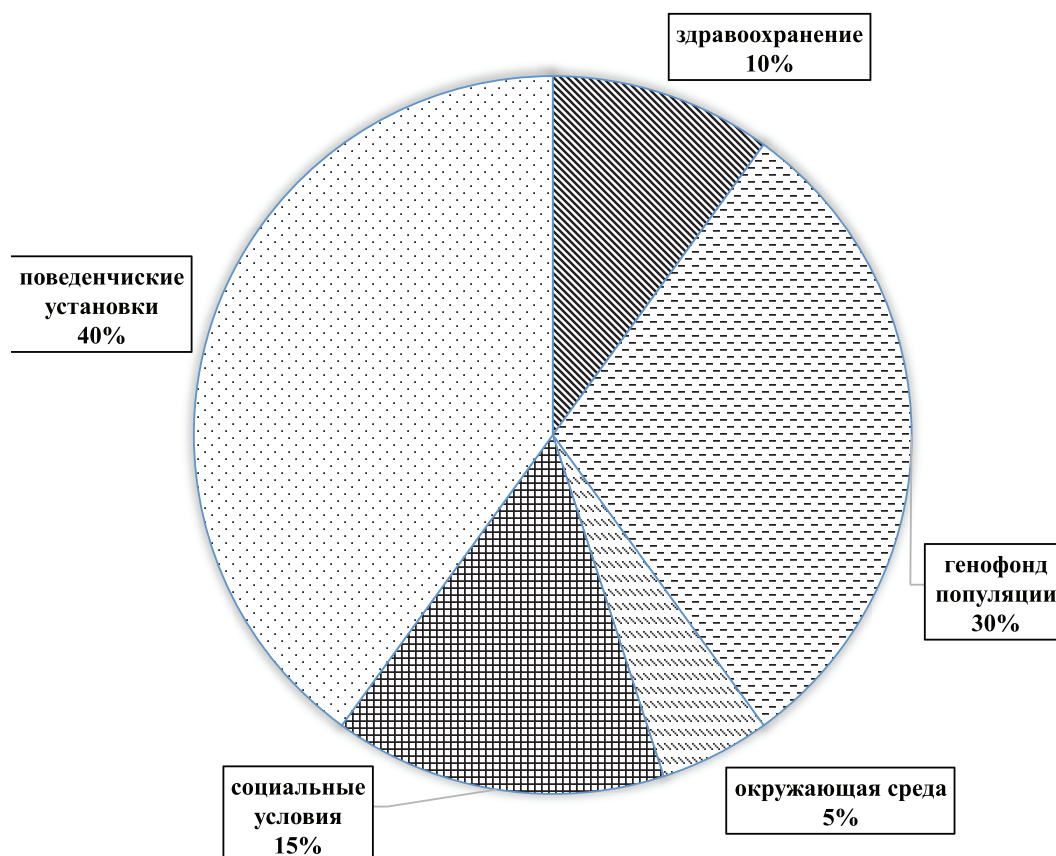
ной логистикой, потоками пациентов, кадровый менеджмент. В поликлиниках приложения ИИ также способны взять на себя некоторые административные функции, такие как: сбор информации для принятия решений о страховом покрытии, составление расписаний, процедуры подготовки запросов на получение разрешений и согласований от вышестоящего руководства.

Хотя мы привыкли рассуждать в первую очередь о применении ИИ в учреждениях здравоохранения, возможности его использования могут оказаться гораздо шире. Это могут быть отдаленно расположенные в сельской местности фельдшерско-акушерские пункты, травматологические пункты, аптеки, санатории, дома отдыха, школы, детские сады, реабилитационные учреждения различных профилей, летние лагеря и проч. Хотя некоторые из них можно считать традиционными средами оказания медицинской помощи, доступность информации в этих средах может существенно отличаться. Аналогично, поиск информации, интерпретация знаний и поддержка принятия решений в таких местах сильно разнятся и поэтому они заслуживают рассмотрения как отдельный тип среды для целей и условий развертывания платформ ИИ. Стоит отметить, что нейросети уже широко используются в немедицинских сферах: для прогнозирования покупок клиентов, управления складскими запасами и т.д. Как только будут решены этические вопросы и проблемы конфиденциальности, то почему бы не начать использовать ИИ для увязки показателей здоровья с рекомендациями по покупке более здоровых продуктов питания для больных гипертензией или диабетом?

Наиважнейшим элементом высококачественной медицинской помощи является укрепление здоровья населения, которое, между прочим, представляет собой непростую задачу для системы здравоохранения, поскольку лишь небольшая часть показателей состояние здоровья нации может напрямую зависеть от уровня оказания медицинской помощи в стране (рис. 3).

Пропаганда здорового образа жизни является важнейшим направлением усилий по укреплению здоровья нации. Часть работы в этом контексте проводится вне регулярных посещений больными лечебных учреждений. ИИ потенциально может помочь в определении приоритетов, управлении объемами и интенсивностью контактов с пациентами, а также в ориентации врачебного внимания на наиболее нуждающихся больных. Одним из компонентов данных инициатив является установление контактов с большим количеством пациентов, что может осуществляться с помощью различных автоматизированных, легко масштабируемых методов, таких как: обмен текстовыми сообщениями и порталы для пациентов. Хорошим примером служит программа по снижению веса лиц с преддиабетом, в ходе которой использовались подключенные к Интернету устройства со специальным приложением, предоставляющим пациентам учебные материалы и отслеживающим параметры их жизнедеятельности. Использование такого ин-





**Рис. 3.** Детерминанты здоровья населения.

струмента тесно коррелировало с нормализацией веса больного в течение нескольких лет. Общение с виртуальным медработником может быть столь же простым, как асинхронный обмен сообщениями с напоминаниями о необходимости сделать прививку от гриппа или прийти на запланированную встречу с использованием защищенных сообщений, автоматических телефонных звонков или электронной почты. Действия более высокого порядка, такие как психосоциальная поддержка или курация лечения хронических заболеваний, могут предполагать использование специальных приложений с голосовыми или видеомодалностями, которые также могут осуществляться посредством телемедицины, что отражает взаимозаменяемость технологических решений [101, 120].

В настоящее время общение врача и пациента все чаще осуществляется в рамках информационного поля (Романова Т.Е. с соавт., Вестник ВШОУЗ 2022; Романова Т.Е. с соавт. Социология медицины 2022, Романова Т.Е. с соавт., 2023). Поскольку виртуальные взаимодействия врач-больной недороги и легко выполнимы по сравнению с личным визитом, вскоре они станут объектом приложения для ИИ. Благодаря усилиям крупных игроков коммерческого здравоохранения, насыщенные большими наборами данных статистические модели стали более активно применяться при прогнозировании среднесрочного и долгосрочного риска для отдельных категорий представляющих целевую аудиторию лиц столкнуться с

неблагоприятными последствиями хронических заболеваний, включая летальные исходы. Указанные прогнозы начали использоваться для определения приоритетности оказания медицинской помощи соответствующим контингентам населения. В подобных ситуациях быстрый анализ информационных массивов является ключевой функцией. И это именно та способность, в которой ИИ несомненно превосходит всех нас [28, 29, 30].

Область приложений, ориентированная на пациентов и осуществляющий уход персонал, объединяет оказание медицинской помощи с общедоступным потребительским оборудованием и программным обеспечением. Она определяется как пространство, в котором приложения и устройства напрямую доступны пациентам и медработникам. Инструменты и программное обеспечение в этой области позволяют пациентам управлять определенной частью своей медицинской помощи и облегчают взаимодействие между пациентами и системой здравоохранения. В частности, смартфоны и мобильные приложения изменили характер контактов врачей с пациентами. Последние стали принимать более активное участие в лечебном процессе; благодаря напоминаниям медработников поведение больных стало гораздо организованнее и рациональнее. Приложения ИИ также обладают несомненным потенциалом повышения качества медицинской помощи, обеспечивая врачам доступ к постоянно обновляющемуся потоку данных от пациентов, который помогает стратифицировать риски, корректировать рекомендации по уходу, предотвращать осложнения хронических заболеваний. Эта тенденция стремительно размывает границы задач, которые традиционно решались в ходе личных встреч.

Кроме того, создаётся весьма многообещающий задел для создания цикла положительной обратной связи, в котором потребности и предпочтения пациентов, зафиксированные посредством приложений ИИ, могут быть учтены другими приложениями, которые в перспективе станут атрибутами системы оказания медицинской помощи. Роль мобильных платформ ИИ в прямом взаимодействии с пациентами по предоставлению рекомендаций в отношении здоровья и лечения продолжает расти параллельно с ролью онлайн покупок предметов ухода, медицинских гаджетов, биологически активных добавок и т.п. Распространение этих приложений будет продолжать расширять и улучшать данные, собранные в ходе традиционной медицинской деятельности. Мобильные приложения все чаще способны связывать различные источники данных и, таким образом, делать подсказки системе здравоохранения путем привязки данных о покупках отдельных пищевых продуктов к показателям здоровья, пройденным за день человеком шагам или статистике использования носимых медицинских устройств. Сбор и представление рекомендаций ИИ с использованием мобильных или настольных платформ имеет решающее значение, поскольку пациенты все чаще участвуют в мероприятиях по самообслуживанию,

поддерживаемых приложениями, доступными на нескольких платформах. Эта технология, скорее всего, будет широко использоваться не только пациентами, но и их семьями и лицами, осуществляющими уход.

К сожалению, прогресс в области использования технологий ИИ, предназначенных для индивидуального управления здоровьем, идёт слишком уж неспешно. В настоящий момент доступно более 300 тыс. приложений для здоровья, и они были загружены без малого 4 млрд. раз. Однако исследований, чтобы определить, улучшили ли они чьё-либо здоровье, практически не было проведено. В обзоре систематических исследований, оценивающих влияние мобильных приложений на здоровье, упомянуто лишь о шести метаанализах, включая в общей сложности 23 рандомизированных исследования. В 11 из 23 исследований были продемонстрированы либо отдельные факты влияния приложений на здоровье, либо суррогатные результаты, но общие доказательства эффективности были признаны неубедительными. Более того, у людей растёт обеспокоенность тем, что эти приложения передают личные данные о состоянии здоровья в неподконтрольные им облачные сервисы, которые рано или поздно оказываются открытыми для злоумышленников (Vyambasuren, et al., 2018) [25, 31, 55].

### **Проблемные аспекты развёртывания платформ ИИ**

Как информация, полученная медицинской наукой, так и технологии ухода за больными, достигли такого уровня сложности, к которому многим врачам сложно адаптироваться (Berdutin V.A., et al., 2023) [51]. Это критически важная проблема, поскольку нарастающая информационная перегрузка приводит не только к недовольству и переработкам медицинского персонала, но и, как неизбежное следствие, к дефектам оказания медицинской помощи.

Регулярный доступ к передовым медицинским знаниям на первый взгляд должен был бы способствовать облегчению лечебно-диагностического процесса, однако на практике всё происходит ровно наоборот. В этой связи ИИ может стать неотъемлемой частью платформ, включающих лучшие научные достижения, не только повышая качество медицинской помощи, но и преобразуя сам характер образовательного процесса врача посредством ментального предоставления нужной клинической информации на этапе принятия решений. Ключевые задачи поддержки должны включать интеллектуальный поиск с извлечением источников информации, настройкой отображения данных, синтезом свежих публикаций об опыте лечения пациентов с аналогичным диагнозом. Гипотетически инструменты ИИ избавят врачей от необходимости делать всё вручную, как это наблюдается по сей день. Поскольку важная информация не будет упускаться из виду, качество и безопасность медицинской деятельности обязательно воз-

растут. Медицинская помощь все чаще и чаще начинает оказываться междисциплинарными бригадами специалистов, в состав которых кроме врачей и медицинских сестёр могут входить клинические фармацевты, генетики, социальные кураторы, разработчики новых медицинских устройств и др. Каждый из них привносит в «общий котёл» специальные навыки и точки зрения в своём квалификационном сегменте, которые дополняют и совершенствуют оказываемую помощь. Поскольку объем данных и информации, доступных для ухода за пациентами, растет в геометрической прогрессии, инновационные решения, основанные на технологиях ИИ, естественным образом станут замечательным подспорьем для работы междисциплинарных команд, предоставляя им экспертные сведения буквально по первому требованию [9, 10].

Алгоритмы ИИ, прогнозирующие будущие результаты на основе существовавших в определенные моменты времени характеристик больших данных, работают с использованием сложных статистических методов и/или машинного обучения. Исходные датасеты могут включать в себя сотни переменных, собранных от сотен тысяч или даже миллионов пациентов. В целом, класс прогнозных приложений ИИ фокусируется на оценке вероятности того или иного исхода для отдельных лиц путем применения пороговых значений риска. В дальнейшем эти больные будут получать то больше, то меньше внимания и ресурсов в плане оказания медицинской помощи, динамического наблюдения или врачебных вмешательств на основе оценки баланса ожидаемого риска, пользы для здоровья, материальных и финансовых затрат. В определенные моменты времени прогнозы могут формироваться как для отдельных людей, так и для групп пациентов, для идентификации лиц/групп с высоким риском неблагоприятного клинического события. Методы прогнозирования индивидуального риска того или иного события известны были давно, но из-за ограниченности доступных данных, например, небольшие выборки или скудные сведения о претензиях без клинической объективизации, точность прогнозов сильно страдала. Появление больших облачных хранилищ данных, извлекаемых из записей в историях болезни, административных баз и других источников, в сочетании с высокопроизводительными вычислениями позволило делать относительно точные прогнозы. В настоящее время существуют отличные инструменты прогнозирования, у которых C-статистика / *C-statistics* площади под кривой *areas under curve* превышает 0,85 или выше. Эти инструменты используются для выявления в амбулаторных условиях пациентов с определенными заболеваниями, которые имеют высокие риски госпитализации и экстренной востребованности неотложной помощи. Этим людям показаны дополнительная координация ухода или госпитализация в профилактических целях из-за нависающей угрозы обострения хронического заболевания (Islam, et al., 2019) [89, 111].

Создание сложных инструментов моделирования наряду с накоплением больших наборов клинических данных привело к



разнообразие рыночного предложения прогностических моделей ИИ. Машинное обучение способно обеспечить большую точность, чем стандартные методики, такие как логистическая регрессия *logistic regression*, вот только, если датасет мал, улучшения могут оказаться совсем незначительными. Такой прирост точности не компенсирует затраты на необходимую для машинного обучения вычислительную инфраструктуру, особенно когда целью является поддержка врачебных решений в реальном времени. Ещё одна проблема заключается в том, что в зависимости от методов, используемых для создания модели машинного обучения, оценка точности прогнозирования модели может оказаться непростой задачей. Если модель обучена просто обеспечивать двоичную классификацию, вероятности не генерируются, и может оказаться невозможным проверить точность прогнозов в диапазоне уровней риска. В таких случаях нелепо строить калибровочные кривые или стратификационные таблицы, что является основополагающим для точности прогноза. Дело в том, что в течение 1 года после разработки модели эффективность калибровки быстро снижается как при вводе, так и при выводе данных. Это влияет на производительность системы *CDS* поддержки принятия решений *clinical decision support*, поскольку чувствительность и специфичность порогового значения модели меняются вместе с падением калибровочной эффективности (Davis et al., 2017) [64].

Смешно надеяться, что врачи станут работать лучше благодаря одним лишь прогнозам. К примеру, управление по охране здоровья ветеранов США разработало оценку потребностей в уходе *CAN care assessment needs*, чтобы помочь врачам выявлять пациентов с высоким риском. Оценка *CAN* отражает клинические и демографические характеристики, которые прогнозировали будущую госпитализацию и смертность для 4 598 408 ветеранов с надёжностью, определяемой областями под кривой *areas under the curve AUC*: для прогнозирования госпитализации 0,84; смерти 0,86; госпитализации и/или смерти 0,82. Первоначальный алгоритм оценки *CAN* имел 90 входных переменных, а текущая версия имела 36 переменных при аналогичной прогностической точности. Эти прогнозы еженедельно предоставлялись более чем 7 тыс. врачам первичного звена по всем категориям пациентов. Однако, судя по статистике использования, только 15% врачей регулярно обращались к этим отчетам, хотя при опросе те, кто использовал эти отчеты, заявили, что в целом считают их точными (Nelson, et al., 2019). Поэтому можно констатировать, что даже если прогностические модели обладают высокой точностью, их развертывание в медицинских учреждениях вряд ли положительно скажется на результатах лечения пациентов, если они не будут тесно связаны с последующими профилактическими мероприятиями, а их рекомендации не будут «с колёс» интегрироваться в производственные процессы учреждений здравоохранения. Для этого врачам пора привыкать к стилю мышления в терминологии пар *прогнозирование-действие*. Под-

водя итог, можно сказать, что, хотя нейросетевые алгоритмы обладают огромным потенциалом для улучшения существующих методов прогнозирования рисков, существуют не решенные вопросы в отношении того, как они интегрируются с потребностями и рабочими процессами медицинских организаций, а также проблемы с обслуживанием платформ ИИ после развертывания. Медицинские работники никогда не должны забывать о негативных последствиях утечки информации об индивидуальной или наследственной предрасположенности пациента к онкологическим, сердечно-сосудистым или психическим заболеваниям [119].

*CDS* охватывает целый ряд приложений ИИ, предназначенных для сообщения врачам новой информации и подсказок при решении сложных задач, в том числе, задач прогнозирования. До сих пор *CDS* в основном применялся в качестве предупреждений, основанных на правилах, например, в виде напоминаний о прививках, или предостережений, использующих формальную логику, основанную на известных индексах риска, которые редко меняются, например, *Framingham risk index*. Более сложные системы были основаны на обширных, основанных на знаниях приложениях, которые помогают в лечении хронических заболеваний, таких как гипертония. Благодаря достижениям в области информатики, включая обработку естественного языка, машинное обучение и многим другим стало возможным отслеживать и моделировать более сложные производственные процессы медицинских учреждений. При условии владения исчерпывающей информацией, платформы ИИ могут адаптироваться к решению самых сложных задач. Подключаясь к другим приложениям, они могут расширять свои поисковые и аналитические возможности, скажем, находить информацию о результатах лечения прошлых пациентов, аналогичных тем, которые в настоящее время нуждаются в терапии. Например, однажды данные электронных историй болезни почти 250 млн пациентов были проанализированы с использованием машинного обучения для определения наиболее эффективных гипогликемических препаратов второго ряда (Vashisht et al., 2018) [115, 157].

Как уже неоднократно упоминалось, одной из клинических областей, в которых ИИ начинает занимать доминирующие позиции, является обработка изображений. В 2017 г. библиография по темам, связанным с компьютерной, магнитно-резонансной томографией, нейрорадиологией и маммографией перевалила рубеж в 800 публикаций. Задачи, для которых нынешний уровень платформ ИИ вполне достаточен, включают: сравнение текущих и предыдущих изображений, определение приоритетов при отслеживании снимков, требующих внимания уже на ранних стадиях заболевания, осуществление высокопроизводительного скрининга, который позволяет рентгенологам сосредоточиться на изображениях, имеющих малозаметные признаки патологических изменений. Со временем вполне вероятно, что интерпретация рутинных изображений будет выполняться исключительно приложениями ИИ. Анализ других типов изображений

с помощью ИИ развивается также в других областях медицины: дерматология, патоморфология, офтальмология. Например, устройства для скрининга диабетической ретинопатии в настоящее время начали использовать не только в крупных академических медицинских центрах, но и в учреждениях первичной медико-санитарной помощи для выявления пациентов, которых следует направить к офтальмологу.

Долгое время предпринимались попытки разработать программы компьютерной диагностики типа Iliad, QMR, Internist, DXplain, однако ни одна из этих систем не получила широкого распространения. Недавно предпринятые компанией IBM Watson усилия также оказались безуспешными. Это объясняется скромным уровнем финансирования в сопоставлении с огромными инвестициями в приложения ИИ для обработки изображений, а также сравнительно большими трудностями диагностики пациентов, чем при интерпретации изображений. Диагностическая информация собирается из многих источников, включая клинические записи, лабораторные анализы, лучевые изображения, сведения из аптек, геномная информация и т.д. Данные часто поступают нестандартизированными, терминологически разнородными и не в цифровом формате. Кроме того, в отличие от визуализируемых изображений, диагностическая деятельность имеет крайне широкий спектр направлений, что делает анализируемую область экспоненциально проблемной. Тем не менее, отдельные учёные пророчат компьютерной диагностике бурное развитие в будущем (Palmer, 2018). С этим созвучна разработка алгоритмов фенотипирования для выявления людей с определенными патологическими состояниями, их классификации по стадии, тяжести и другим нозологическим характеристикам с использованием данных электронных медицинских карт. В настоящее время не существует общего стандартизированного, структурированного, вычислимого формата для алгоритмизации фенотипирования, но семантические подходы активно исследуются, обещая более точную идентификацию, чем банальное использование диагностических кодов. Геномная привязка фенотипа с помощью инструментов ИИ значительно расширяет возможности диагностики и понимания молекулярной основы заболеваний. Со временем достижения в этой сфере начнут стимулировать поиск новых способов лечения [127].

### **Организационное управление и эксплуатация платформ ИИ**

Маркетологи, занимающиеся продвижением инструментов ИИ, раз от раза наращивают усилия, делая громкие заявления о потрясающих клинических результатах и невероятном повышении эффективности оказания медицинской помощи, которые хотя и могут быть частично обоснованными, но по сути представляют собой беззастенчивые манипуляции. В этом разделе мы затронем вопросы структуры оценок принятия решений по внедрению платформ ИИ, которые включают в себя соображения, касающиеся ор-

ганизационного управления и технической эксплуатации, а также некоторые клинические аспекты, критически важные для успешной реализации.

Обратимся к истории: более 10 лет назад американская некоммерческая неправительственная организация Национальная академия медицины *NAM National Academy of Medicine*, известная до 2015 г. как Институт медицины *IoM Institute of Medicine*, предложила новое понимание системы здравоохранения, как динамичной системы, которая не просто предоставляет медицинскую помощь традиционным способом, основываясь на клинических руководствах и профессиональных нормах, но постоянно оценивает и совершенствует себя, используя возможности информационных технологий. Таким образом, система здравоохранения должна представлять собой систему с жёстко встроенным в её операционную модель алгоритмом постоянного обучения. Исходя из этой концепции, огромное количество данных, генерируемых в процессе оказания медицинской помощи, естественным образом становятся доступны для включения в деятельность по постоянному совершенствованию самой системы здравоохранения, а также для исследований, направленных на улучшение качества медицинской помощи в целом. Таким образом, оказание медицинской помощи каждому отдельному пациенту начинает приносить фактическую пользу обширной аудитории аналогичных больных. Чтобы подсветить контекст того, почему обучающаяся система здравоохранения *LHS learning health system* имеет судьбоносное значение для развития медицинского ИИ, мы решили привести здесь некоторые рекомендации из отчета *NAM*, увязав их с тем, как ИИ можно рассматривать в рамках *LHS*.

К разработке платформ ИИ рекомендуется активнее привлекать пациентов и их семей. Во-первых, людей необходимо доходчиво информировать о роли приложений ИИ в оказании им медицинской помощи. Во-вторых, всячески стимулировать участие пациентов к совместному с ИИ принятию решений по лечению. Хотя меры по налаживанию взаимодействия при принятии решений еще не принесли положительных результатов, но их перспективность ни у кого не вызывает никаких сомнений, поскольку многие приложения находятся на очень ранней стадии разработки. Внедрение нейросетевых технологий предоставляет хорошую возможность для продвижения идеи *LHS*, что требует наличия развитой цифровой инфраструктуры. Одна крупная больничная сеть США как-то объявила об амбициозных планах по включению ИИ в каждое взаимодействие с пациентами, во все рабочие и административные процессы, чтобы стимулировать улучшение качества медицинских услуг и снижения их себестоимости. Многие другие медицинские компании хотели бы поддержать эту инициативу, хотя понимают, она несколько преждевременна. Технологии ИИ и цифровая инфраструктура здравоохранения должны созреть до того уровня, чтобы генерируемые данные годились не только для



оценки производительности самих нейросетевых платформ, но и для детального понимания всех нюансов функционирования системы здравоохранения в целом, а также отношения к ней обслуживаемых контингентов населения. В этой связи крайне важно, чтобы жизненный цикл использования ИИ находился под непрерывным контролем уполномоченных государственными органами управления ответственных лиц.

Их главной задачей является управление данными на основе методологии, которую медицинские организации применяют для обеспечения соответствия обрабатываемых данных установленным стандартам, нормам и правилам. Учитывая высокие требования ИИ к качеству данных, управление данными имеет решающее значение и распространяется на вопросы, связанные с курированием данных и конфиденциальностью. Такие приложения, как *SMART on FHIR / Substitutable Medical Apps, Reusable Technology on Fast Healthcare Interoperability Resource*, т.е. заменяемые медицинские приложения, многоразовые технологии на ресурсе быстрого взаимодействия в здравоохранении, являются новым благом из области ИИ. Вместе с тем, они могут усугубить проблемы, связанные с передачей информации из электронных медицинских записей во внешние системы, что осложняет обеспечение безопасности данных. Кроме того, общественность бьёт тревогу из-за нерешенности многих этических вопросов, связанных с утечками конфиденциальных данных. В литературе есть целая серия публикаций, посвященных правам и предпочтениям пациентов, а также тому, как *LHS* превращается в моральный императив прогрессивного человечества. Там утверждается, что люди готовы делиться информацией в целях улучшения общественного здоровья и облегчения взаимодействия с медицинским персоналом (Morain, et al., 2018; Wicks, et al., 2018). Другими словами, эти публикации предполагают, что медицинские данные следует воспринимать как общественное благо. Для организаторов здравоохранения можно опираться на такого рода публикации, апробируя инновационные управленческие подходы, которые бы гарантировали, что данные о здоровье населения использовались исключительно в интересах больных. Не забывайте, что не все пациенты и/или их близкие обладают адекватным уровнем культуры и грамотности для понимания того, насколько правильно / законно персональные данные могут быть использованы или монетизированы [115, 163].

Во главе угла эксплуатационной стратегии платформ ИИ стоит системный подход, общепризнанной фундаментальной концепцией которого является цикл «планируй-делай-исследуй-действуй» *plan-do-study-act PDSA*, предложенный отцом современного подхода улучшения качества W. Deming. Цикл *PDSA* опирается на активное участие сотрудников, вовлеченных в работу, детальное понимание рабочих процессов и тщательную постоянную оценку реализации, которая дает информацию для итеративных корректировок. Новые методы повышения качества, появившиеся со времен Деминга,

представляют собой вариации или усовершенствования этого подхода. Однако слишком часто усилия по улучшению качества, в том числе использующие инструментарий *бережливой медицины*, терпят неудачу, потому что они слишком узко сфокусированы на конкретной задаче или наборе задач с использованием неадекватных и/или нерелевантных показателей, а также без должного учета системного окружения, т.е. той среды, в которой будут происходить изменения (Бердугин В.А. с соавт., 2021) [5, 8].

Развертывание и обслуживание платформ ИИ должно осуществляться строго по изложенным принципам. Иначе, если моральное состояние медицинского коллектива изначально находилось на низком уровне, то интеграция алгоритма ИИ для совершенствования уже налаженных производственных процессов и методик принятия решений могут рассматриваться персоналом как трудозатратное и избыточное дополнение к и без того непопулярной корпоративной информационной системе учреждения здравоохранения. Более того, существует ряд характерных для ИИ проблем, наличие которых вполне объяснимо, а происхождение других не очень понятно. В этой связи мы утверждаем, что одних классических подходов к обеспечению качества работы ИИ абсолютно недостаточно. Этот тезис в первую очередь касается намерения осуществлять крупномасштабные проекты. Нам на выручку приходит специально созданная технология внедрения, которая декларирует то, что организации смогут проводить систематические изменения только, признавая их сложность. В целом, эта технология делает упор на разумные подходы к управлению изменениями, которые, несомненно, применимы в деле реализации платформ ИИ. Упомянутые концепции управления изменениями объединяют базовые элементы, которые в принципе должны быть знакомы даже начинающему руководителю, поскольку они регулярно упоминаются в литературе по менеджменту. Эти концепции и способы их адаптации к конкретной задаче по реализации платформы ИИ включены в таблицу 2. Наконец следует признать, что даже когда руководство медицинской организации вполне компетентно, процесс реализации платформы ИИ может идти совсем не так, как планировалось или ожидалось, особенно если речь идет о крупном учреждении здравоохранения. Ну что ж, концепции того, как успешно достигать желаемых результатов производственной деятельности, продолжают развиваться, всё больше опираясь на изощренные организационные инновации, которые при аккуратном применении способны принести много пользы общественному здравоохранению.

Характер усилий государства и общества по совершенствованию системы здравоохранения исторически носит циклично-итеративный характер (рис. 4). Обычно процесс начинается с осознания некоей проблемы и необходимости её решения. За проблему принимают препятствие/барьер, мешающий поступательному развитию медицинской организации или клиническому направлению, а также разочарование врачей и пациентов из-за

Таблица 2.

**Релевантность аспектов ИИ ключевым концепциям и задачам реализации нейросетевых платформ.**

Задача или концепция реализации	Релевантные аспекты ИИ
Определение клинической или административной проблемы, которую необходимо решить.	Рассмотрение проблемы, которую необходимо решить, должно предшествовать выбору и внедрению конкретных технологий, таких как системы ИИ, и отличаться от них.
Оценка готовности мед. организации к изменениям включает опрос заинтересованных сторон, как это предусматривается концепцией «Продвижение действий по внедрению исследований в здравоохранение» <i>Promoting Action on Research Implementation in Health Services framework</i>	Важно привлечь к участию врачей, специалистов в области информационных технологий и организаторов здравоохранения. Стейкхолдеры необходимы для эффективного планирования организационной подготовки к имплементации платформ ИИ на местах.
Достижение консенсуса среди стейкхолдеров о том, что проблема весьма актуальна, и предоставление убедительных свидетельств того, что предлагаемое решение окажется эффективным.	
Без излишней жесткости и нажима применять стандартные и хорошо знакомые персоналу / пациентам организационные подходы, подбирая наиболее целесообразные методики адаптации.	Разработка и принятие стандартных подходов к подготовке датасетов, разработке моделей ИИ и оценке показателей производительности наряду с использованием дружелюбных пользовательских интерфейсов и обучения персонала владению новыми инструментами.
Оценка влияния ИИ на производственный процесс, результаты лечения пациентов и организационную эффективность учреждения здравоохранения в целом.	Необходимо четко сформулировать ценностное предложение, сравнив текущую ситуацию в медицинской организации с масштабом ожидаемых улучшений благодаря эксплуатации ИИ.

Задача или концепция реализации	Релевантные аспекты ИИ
Обеспечение твердой приверженности проекту со стороны высшего руководства мед. организации.	Административный блок управления проектом должен включать: высший менеджмент мед. организации, руководителей лечебно-диагностических подразделений, инженеров по обслуживанию компьютерной техники, специалистов в области экономики и финансов, команду экспертов по разработке стратегий, директив управления и определения организационных приоритетов.
Выявление идеологически заряженных сотрудников и сильных лидеров среди клиницистов.	Каждая внедренная на практике нейросетевая платформа нуждается в опытных клинических владельцах, которые могут мастерски использовать возможности ИИ, заранее предвидя и предупреждая сбои работы алгоритма.
Привлечение стейкхолдеров к разработке плана реализации платформы ИИ, который был бы осуществим и приемлем для всех пользователей, а также определение размера компенсаций, если потребуется дополнительная работа со стороны пользователей.	Инструменты ИИ должны быть спроектированы и реализованы с четкой ориентацией на пользователей, имеющей целью снижение рабочей нагрузки, сбережение времени и когнитивных усилий. Поэтому и нужны подробные планы, отражающие будущие изменения в рабочем процессе, информационной логистике, установку нового оборудования и т. д.
Обеспечение надлежащего обучения и технической поддержки во время реализации проекта.	Внедрение инструментов ИИ в учреждениях здравоохранения должно сочетаться с образовательными инициативами, с круглосуточной технической и клинической поддержкой пользователей, которая в идеале должна быть способна с ходу устранять возникающие проблемы.



Задача или концепция реализации	Релевантные аспекты ИИ
Устранение проблемы избыточной сложности существующих информационных систем из-за наличия навязчивой функции «выбора».	Использование интуитивно понятных пользователям интерфейсов и оптимизация рабочих процедур уже на этапе проектирования.
Четкое определение этапов, показателей и результатов для оценки успешности проекта.	Целевое состояние платформы ИИ должно быть четко сформулировано и описано в измеряемых величинах. Проектом должны быть предусмотрены автоматизированные процедуры сбора, анализа и своевременного предоставления требуемой информации.
Подведение итогов после очередного принятия мер, которые будут служить основой для дальнейших усилий по реализации.	Изучение специальной литературы по взаимодействию человека и компьютера для оценки восприятия пользователями препятствий и уроков, извлеченных в ходе внедрения инструментов ИИ.

невозможности улучшить производственные процессы. Для медицинского менеджмента первоочередной задачей является определение масштабов и остроты проблемы, убедиться в том, что она не носит частный характер и что не существует более простых подходов к её решению, чем реализация крупного информационного проекта. Начинать нужно с описания существующих рабочих процедур, что обычно предусматривает углубленные интервью с персоналом и прямое наблюдение, которые помогают составить подробные блок-схемы ситуации на местах (Бердугин В.А. с соавт., 2020) [7].

Потом нужно определить желаемое состояние результата, и все возможные варианты его достижения следует рассмотреть и сравнить. На каждом последующем этапе процесса реализации платформы следует запрашивать обратную связь с мнениями стейкхолдеров, таких как: пациенты, медицинский персонал, авторитетные представители общественности. Хотя тема ИИ в настоящее время являются крайне популярной и всюду рекламируется, хотим напомнить нашим любознательным читателям, что она приближается к пику цикла хайпа Гартнера *Gartner Hype Cycle*, и некоторые нейросетевые приложения по совокупности неприятностей и убытков рискуют превзойти реально достижимую пользу. Таким образом, вполне вероятно, что ранние реализации платформ ИИ по затратам покажутся великанами на



**Рис. 4.** Жизненный цикл разработки приложений искусственного интеллекта.

фоне полученной прибыли. А вот более осторожный медицинский менеджмент, готовый отложить инвестиции до тех пор, пока не удостоверится в зрелости продукта, хоть и не будет числиться в передовиках, зато сохранит финансовую устойчивость. В конечном итоге каждому руководителю медицинской организации рано или поздно придётся заниматься оценкой полезности эксплуатации платформ ИИ. При рассмотрении целесообразности развертывания того или иного приложения ИИ для прогнозирования возможно, что в определенных ситуациях, учитывая стоимость, логистическую сложность и степень эффективности алгоритма, может не оказаться подходящих рабочих зон, в которых пара прогнозирование-действие будет иметь клиническую ценность.

### **Клиническая валидация и мониторинг инструментов медицинского ИИ**

Клиническую валидацию инструментов медицинского ИИ следует рассматривать отдельно от технической валидации, многие аспекты которой были описаны ранее. Для ИИ клиническая валидация имеет 2 ключевые оси:

1. Применение традиционной медицинской иерархии доказательств для стимуляции приятия и постоянного использования платформ ИИ. Результаты пилотных проектов классифицируются как самый низкий уровень доказательств, за которыми следуют оценки работы алгоритма по итогам наблюдений с поправкой на

риск. Результаты клинических испытаний располагаются на вершине иерархической пирамиды.

2. Соответствие цели ИИ желаемому клиническому состоянию. Одна лишь демонстрация высокой прогностической точности не обязательно гарантирует улучшение результатов лечения, если не предпринимать целенаправленных действий по их достижению, а также если алгоритм прогнозирует изменение производственного процесса, которое никак не влияет на повышение качества медицинской помощи. Таким образом, критически важно определить пары прогнозирование-действие. Причем действия должны заключаться не просто в накоплении знаний и опыта, а должны состоять в конкретных мероприятиях, непосредственно ведущих к лучшим показателям лечебно-диагностического процесса.

Учитывая необычность технологий ИИ, неубедительность доказательств необходимости его безоговорочного использования, рыхлость его нормативной базы и полную зависимость нейросетевых платформ от нюансов во вводимых клинических данных, постоянный мониторинг показателей работы ИИ в учреждениях здравоохранения приобретает решающее значение в обеспечении его безопасного и эффективного использования. Основой для мониторинга за парой прогнозирование-действие должна служить оценка следующих факторов:

- как часто инструмент ИИ востребован и используется при оказании медицинских услуг;
- как часто рекомендации ИИ принимаются во внимание либо игнорируются с указанием причин;
- насколько удобны настройки конфиденциальности, вызовов журнала доступа к данным и интерфейса прикладного программирования;
- как измеряются клиническая безопасность и польза для пациента. Оптимально это делать в форме согласованных результатов;
- наличие организационных метрик, относящихся к производственному процессу или бэк-офису ИИ;
- отсутствие неточных или неоправданно отвлеченных рекомендаций и несвоевременных или неверных подсказок;
- обязательность записей о текущих работах по техническому обслуживанию, в том числе, запросов на пересмотр данных;
- сохранение производительности модели без потери мощности из-за нововведений в документации.

Сложность и масштаб локальной оценки и мониторинга могут неизбежно различаться в зависимости от манеры внедрения ИИ в производственный процесс, клинической ситуации и типа предоставляемой *CDS clinical decision support*, поскольку они, в свою очередь, определяют приписываемый нейросети клинический риск.

Фреймворк Международного форума регуляторов медицинских устройств *IMDRF International Medical Device Regulators Forum* для оценки риска *SaMD software as a medical device* / программное

обеспечение как медицинское устройство является потенциально полезным инструментом для разработки стратегий мониторинга, адаптированных к уровню потенциального риска клинических ситуаций. IMDRF фокусируется на характеристиках места оказания помощи, например, отделение интенсивной терапии, типе предлагаемого решения, скажем, состояние, угрожающее жизни больного, типе предоставляемой поддержки принятия решений, например, резко всплывающее оповещение. В целом, потенциальная потребность в мониторинге возрастает вместе с остротой клинической ситуации и необходимостью моментального принятия решений, а также по мере того, как снижается участие разработчиков нейросети в её поддержке и выявлении ошибок.

Перспективы высокорисковых платформ медицинского ИИ целиком определяются степенью их безопасности и эффективности, при этом допустимо смотреть сквозь пальцы на прочие важные показатели, такие как: вызовы *API application programming interface* с информацией об опыте пользователя. Чтобы читателю стало более понятно, развернем данную мысль. Интерфейсы прикладного программирования *API* нужны для взаимодействия одной компьютерной программы с другой. Вызовы *API* – это среда, с помощью которой программы взаимодействуют друг с другом; запрос *API* выглядит как сообщение, отправляемое на сервер с просьбой к *API* о предоставлении информационной услуги. Поэтому высокорисковые алгоритмы нуждаются в тщательнейшей проверке надёжности и регулируемости, а также в постоянном мониторинге в процессе развертывания и последующей эксплуатации. Напротив, оценка ориентированных преимущественно на административные функции низкорисковых платформ ИИ должна быть, прежде всего, количественной и сфокусированной на показателях медицинской помощи, проявляющих изнанку работы алгоритма с положительной или отрицательной стороны. Развертывание платформ ИИ рекомендуется по возможности проводить после рандомизированных контролируемых испытаний или А/В-тестирования. Мощный толчок развитию инструментов медицинского ИИ могла бы дать программа крупномасштабных испытаний подобного рода в нескольких медицинских центрах федерального уровня, однако для организации локального мониторинга и настройки формуляров ИИ это будет совершенно излишне.

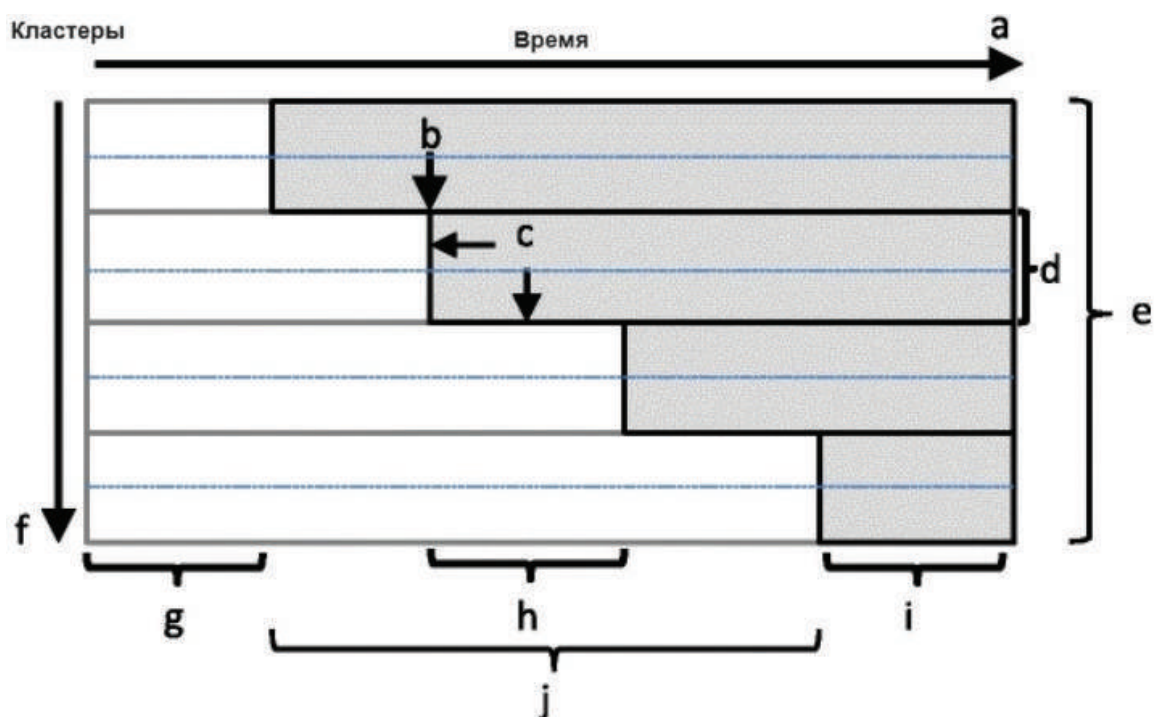
Для частных узкоспециализированных проектов из-за их высокой стоимости, сложностей с выполнением, временными и прочими ограничениями проведение рандомизированного исследования будет затеей экономически нецелесообразной и авантюрной. Для таких условий больше подходят квазиэкспериментальные подходы, такие как: ступенчато-клиновые дизайны *stepped-wedge designs* или даже тщательно скорректированные ретроспективные когортные исследования. Здесь нам опять нужно объяснить читателю более подробно, что это такое. В *stepped-wedge designs* кластеры рандомизируются на несколько групп, которые определяют, когда



начнется итерация. Кластеры начинают испытание при некоем контрольном условии. Группы кластеров переходят в состояние итерации по ступенчатому графику, но все группы будут задействованы до окончания испытания. Ступенчато-клиновые дизайны нужны для проверки гипотезы об выполнимости проекта, так как в некоторых случаях непрактично или вовсе невозможно вовлечь в исследование одновременно всех участников. Ступенчато-клиновая рандомизация решает эту проблему, вводя группы кластеров в итерацию в разное время (рис. 5).

Характеристики и терминология рандомизированных контролируемых испытаний с использованием ступенчатого кластера, где затененные области указывают на итерации, а незатененные области являются контрольными: а. продолжительность испытания, b. точка перехода, c. 2-й шаг, d. количество кластеров во 2-й группе, e. количество групп, f. общее количество кластеров, g. период до развертывания, h. время между 2-й и 3-й точками перехода, i. период после развертывания, j. период развертывания. Результаты мониторинга после внедрения позволяют проводить тщательную оценку, подобную той, какую осуществляют корпоративные компьютерные системы, регулярно отслеживая расход и заказ медикаментов в лечебных отделениях больниц. С учетом вышесказанного наивно было бы думать, что чрезмерно оптимистичная оценка производительности платформы ИИ может оказаться хоть сколько-нибудь полезной для системы здравоохранения [159, 160].

В исследовании 2018 г. Zech с коллегами оценивали способность алгоритма ИИ обобщать получаемые из нескольких кли-



**Рис. 5.** Разработка ступенчато-клинового дизайна с эффектами переноса и рандомизацией.

ник визуальные данные и, таким образом, диагностировать пневмонию по рентгенограммам пациентов. Оказалось, что производительность нейросети при тестировании на внешних наборах данных значительно падала. Приложение ИИ не смогло обобщить данные по больницам из-за различий между данными обучения и данными оценки. Это уже ранее упомянутая нами хорошо известная, но часто игнорируемая разработчиками проблема сдвига набора данных. Этот случай доказывает, что даже незначительные различия в распространенности заболевания, в частности, пневмонии между популяциями приводят к снижению производительности нейросети. Сдвиг наборов данных может вызываться различиями между производителями сканеров или типами сканеров, которые могут быть как портативными, так и стационарными. Тот же эффект наблюдается при системной несопоставимости анализируемых рентгенограмм: инверсия цветовой схемы или наличие на изображении какой-то надписи. По хорошему платформа ИИ может быть обучена лишь точно определять, из какой конкретно больницы поступило изображение; и лучше не ориентироваться на её заключения, сделанные на основании обобщенных данных из разных лечебных учреждений (Zech, et al., 2018) [168].

Эффективность медицинского ИИ может со временем ухудшаться, особенно когда в учреждении здравоохранения меняются клинические рекомендации, протоколы ведения больных или демография популяции. Допустим, в конкретной больнице принято назначались анализ на содержание лактата в крови только при подозрении на сепсис. Нейросеть для прогнозирования сепсиса, обученная на данных из этой больницы, будет неполноценной потому, что факт измерения лактата связан исключительно с сепсисом, а не с иными причинами. Однако если клинический протокол изменится и измерения лактата будут назначаться чаще и по иному поводу, то связь, закрепленная алгоритмом, утратит абсолютный характер. Тогда не только «если А, то только В», а может быть уже «если А, то В, С, D, E...». В качестве альтернативы, если популяционный состав населения изменится, включив большой контингент деклассированных личностей, употребляющих наркотики, то обнаружение у пациента повышенных показателей лактата крови может стать рядовым явлением, и, соответственно, факт назначения анализа на лактат для ИИ уже не будет играть большой роли. В обоих описанных вариантах изменений обучающих датасетов производительность нейросети, вероятно, сильно ухудшится, что приведет к увеличению ложноположительных предупреждений о сепсисе.

Описанное объясняет необходимость скрупулёзной оценки надежности алгоритма ИИ. Ключевым методом измерения его надежности являются проверки на нескольких датасетах. Классически обученные алгоритмы склонны тиражировать артефакты, специфичные для источника данных или для самого обучающего датасета. Существует

множество мало уловимых причин, по которым в наборах реальных данных может возникать смещение. Проверка с использованием внешних наборов данных покажет снижение производительности для платформ, обученных по нескольким принципиально несовместимым шаблонам, которые не могут быть обобщены. Иные факторы, которые могут снижать точность прогнозирования нейросети: страховое покрытие, дискриминационные практики или ограничения ресурсов. В целом, когда речь идёт о неточно измеренных параметрах, небрежно проведенной классификации результатов или маркировки случаев, машинное обучение *machine learning* неизбежно приводит к утечке причинно-следственной связи *causality leakage* и утечке маркировки *label leakage*. Примером утечки причинно-следственной связи в клинических условиях может служить ситуация, когда врач, подозревая наличие у пациента заболевания, чтобы подстраховаться заказывает углубленное обследование, а ИИ использует сам факт заказа для генерации оповещения о якобы собственном решении предупредить врача о проблеме. Таким образом, обратная связь приобретает неконструктивный характер и может привести к гипердиагностике из-за того, что врач полагался на авторитет ИИ, а нейросеть на мнение врача: «кукушка хвалит петуха за то, что хвалит тот кукушку». Утечка меток происходит, когда информация о результате целевой задачи просачивается обратно в функции алгоритма, используемые для генерации результатов.

Несмотря на то, что внешняя проверка проявляет степень надежности алгоритма, она всё же носит реактивный характер, поскольку различия между средами обучения и оценки обнаруживаются постфактум, т.е. по мере ухудшения производительности платформы ИИ. Чтобы избежать сбоев, лучше обнаруживать и предотвращать проблемы проактивно, т.е. ещё до начала обучения. На сегодняшний день уже разработаны проактивные методы обучения, которые настраивают нейросеть делать прогнозы, инвариантные к ожидаемым сдвигам в популяциях или наборах данных. В нашем примере с лактатом в модель ИИ может быть инсталлирован предиктивный алгоритм, который невосприимчив к сдвигам данных. Тут требуется корректировка с учетом смещения, что возможно только когда данные соответствуют определенным требованиям к качеству. Тогда сдвиги можно как бы загодя предвидеть и, поэтому указать об этом в сопроводительной документации. Использование предиктивных алгоритмов требует осторожности и тонкой настройки производительности, дабы не повредить надежности работы нейросети при развертывании в новых условиях. Помимо проактивного обучения, этот подход также важен для понимания восприимчивости определенной модели ИИ к сдвигам данных. Особенно нужны такого рода инструменты на начальных этапах разработки алгоритма.

Кроме мониторинга общих показателей производительности необходимо оценивать производительность моделей ИИ на ключевых подгруппах пациентов. Этот прием помогает дополнительно выявлять локусы уязвимости, поскольку высокая средняя производи-

тельность отнюдь не является гарантией удовлетворительной производительности в подгруппах. Доскональное изучение стратифицированной производительности помогает выявлять подгруппы, в которых алгоритм работает плохо и, следовательно, на данном участке риск ошибки будет велик. Обнаружение таких локусов ненадёжности путём ручного аудита можно считать профилактикой аномалий платформ ИИ. Раньше возможности по оценке неопределённости алгоритмов были не столь широкими как сейчас, когда появились инструменты-оболочки, способные проверять отдельные модели черного ящика (Schulam & Saria, 2019) [143].

Итак, можно констатировать, что эффективность платформ ИИ неизбежно будет ухудшаться, когда они начнут применяться в обстановке, имеющей системные отличия от той, в которой они обучались. Ухудшение эффективности с течением времени может также происходить в пределах одной и той же организации здравоохранения, поскольку изменяется окружающая среда, а с ней и популяционные характеристики проходящих лечение пациентов. Кроме того, само по себе развёртывание платформ ИИ способно вызывать революционные преобразования в процессах оказания медицинской помощи. Существует ряд подходов, используемых для учета изменений в исходных данных, которые ведут свою родословную из традиционных статистических методов по расчету рисков. Они варьируют от полной регенерации платформ ИИ на периодической основе до их повторной калибровки с использованием сложных методик. Однако по-прежнему остаются открытыми вопросы относительно того, как часто обновлять, какой объем и типы данных необходимы для надежного поддержания производительности и как масштабировать действия по мониторингу и обновлению того, что в будущем будет представлять собой массивный стек алгоритмов. Некоторые из грядущих рисков аналогичны рискам систем, в которых были успешно реализованы инструменты *CDS clinical decision support* / поддержки клинических решений на основе булевых правил. Однако постоянное добавление напоминаний от *CDS* в электронные истории болезни на основе различных клинических руководств делает их функционирование неустойчивым и неудобным. Таким образом, без грамотной автоматизации, соответствующего масштабирования и упорядочивания знаний систему управления модулем *CDS* ждут неудовольствие пользователей и серьезные испытания на прочность.

## **Интерпретируемость рекомендаций ИИ**

Отсутствие прозрачности и интерпретируемости рекомендаций искусственного интеллекта, является одной из давно известных проблем, ограничивающих его широкое использование в медицине. Туманные и часто псевдонаучные представления многих врачей о нейросетевых технологиях порождают бесконечные вопросы о степени обоснованности и достоверности результатов, которую



платформы ИИ могут гарантировать стейкхолдерам, принимающим решения в отношении лечения или страховой ответственности. На этом фоне совсем не удивительным явлением выглядит модная тенденция задаваться вопросом, можно ли использовать не только прямые указания ИИ, а лишь некоторые его ассоциации для определения конкретных медицинских мероприятий, которые следует предпринять, а также помогает ли это в обнаружении новых неожиданных патофизиологических феноменов. На самом деле эсхатологически упоминаемая здесь проблематика отнюдь не проста, поскольку она апеллирует к моделям прогнозирования рисков, которые, с одной стороны, проверяемы с позиции точности прогноза, а, с другой стороны, для которых включение элементов данных может не основываться на биологических связях, что уводит нас далеко в сторону от клинических интерпретаций.

Требования к интерпретируемости, вероятно, будут определяться рядом факторов, включая:

1. юридическую ответственность медицинского персонала, нормативно-законодательную базу с рекомендациями по использованию ИИ в здравоохранении;

2. растущее доверие медицинской общественности к технологиям ИИ;

3. влияние медицинского ИИ на взаимодействие врачей и пациентов в плане принятия его участия в лечебно-диагностическом процессе;

4. позитивные ожидания населения относительно безопасности и эффективности нейросетевых платформ.

Популярным направлением инноваций является синтез и отображение того, где и как ИИ интерпретирует данные для передачи конечным пользователям. Сюда относится метод параллельных моделей *parallel models*, где одна используется для основных вычислений, а другая для интерпретации. Ещё есть методы, использующие новые графические дисплеи и инструменты обнаружения данных, которые работают вместе с ИИ, чтобы обучать и помогать пользователям в учреждениях здравоохранения.

Парадокс машинного обучения состоит в создании основанных на признаках алгоритмов, которые очень нелегко интерпретировать. Однако при отсутствии прозрачности *ML*, тем не менее, обеспечивает уверенный контроль и строгие стандарты производительности. Мы можем не видеть и не понимать, как работает большинство компонентов алгоритма, но если его рекомендации абсолютно точны и осуществляется мониторинг производительности платформы ИИ, то мы можем спокойно доверять ей выполнение поставленных задач. Львиная доля медицинских приложений ИИ не заточена на оценку потребности конечных пользователей в уровне интерпретируемости. Хотя никто не будет спорить в том, что сфера здравоохранения нуждается в самых строгих критериях транспарентности нейросетевых моделей, на практике же точность предсказаний ИИ оказывается гораздо предпочтительнее прозрачности. Независимо

от качественного уровня интерпретируемости выходных данных ИИ, на этапе разработки следует обращать внимание на желания пользователей. Имея это в виду, есть несколько ключевых факторов, связанных с развёртыванием платформ ИИ, которые алгоритмически непрозрачны, но очень востребованы и отдельными врачами, и системой здравоохранения в целом.

Нам кажется, что потребность в алгоритмической прозрачности на местном уровне сильно преувеличена. Тем не менее, для сопоставления с критериями включения/исключения результатов испытаний медицинских приложений дескрипторы того, как данные были собраны и агрегированы, всё же необходимы. У моделей, которые устанавливают связи между наблюдаемыми данными и результатом, т.е. предназначены только для прогнозирования, нужно исключать из анализа популяционные контингенты, где ИИ неприменим. Это позволит избежать внесения в модель ложных данных, связанных с социально-экономическими предвзятостями или ошибками в документах, и поможет принимать обоснованные решения о желаемой точности использования алгоритма в конкретных ситуациях. Производители и добросовестные пользователи программного обеспечения должны заранее убеждаться в том, что анализируемые данные были собраны в условиях, аналогичных или очень близких к тем, в которых производилась разработка приложения. Возможность обобщения данных является особенно важным критерием при оценке пригодности инструмента ИИ. Если данные были получены за пределами целевой среды, то, скорее всего, они будут истолкованы превратно. Попытки обобщать несопоставимые данные приводят к некорректной работе моделей ИИ. Мы категорически не советуем экстраполировать результаты, полученные из данных одной узкой популяции пациентов, на популяции, проживающие в других условиях и в иной временной период.

В моделях ИИ, предназначенных для терапевтических целей и составления лечебных рекомендаций, подразумевается наличие жёстко встроенных причинно-следственных связей, которые требуют очень высокого уровня проверки корректности их функционирования. Дело в том, что результаты интерпретации могут быть неточны из-за упомянутой выше изменчивости социально-экономических и системных условий, а также из-за разночтений в документации и временных смещений. К сожалению, инструменты ИИ не способны учитывать такие смещения в автономном режиме. Поэтому в обозримом будущем этот класс инструментов ИИ потребует проактивного изучения перед локальным развёртыванием, и, возможно, повторяющихся калибровок и переоценки при расширении за пределы исходной среды.

Реальная ситуация в отечественном здравоохранении говорит о том, что развёртывание технологий ИИ требует от медицинского менеджмента не только твёрдой решимости, но и немалых ресурсов. Преодоление столь высоко поднятой планки под силу только

крупным хорошо обеспеченным учреждениям здравоохранения, в то время как мелкие и средние учреждения, которые обслуживают непривилегированный класс населения, в основном состоящий из лиц с крайне низкими доходами, пока не смогут позволить себе такую дорогостоящую технологию. На заре медицинского ИИ, когда он находился, по сути, в зачаточном состоянии, это могло быть не так уж критично. Однако по мере совершенствования нейросетей цифровой разрыв может увеличить отставание, которое малые игроки уже не смогут преодолеть. А ведь им инструменты ИИ необходимы в первую очередь, так как именно их медицинская помощь срочно нуждается в повышении качества и эффективности. Дополнительный источник цифрового неравенства заключается в том, что платформы ИИ, как правило, разрабатываются в однородной среде, где отсутствуют пациенты с резкими различиями в социально-экономическом и культурном уровнях, что, естественно, приводит к падению производительности алгоритмов в смешанных группах, состоящих из случайных людей. Поэтому с самого начала разработки модели ИИ алгоритм должен обучаться на основе данных, собранных от максимально разнообразных контингентов населения. Только тогда он будет приспособлен беспристрастно анализировать информацию, поступающую как от состоятельных, так и от бедных пациентов.

Поскольку приложения медицинского ИИ становятся все более доступными, а их маркетинг, прямо на глазах агрессивно усиливается, сопровождаясь промоакциями с заявлениями о великолепных результатах ИИ в диагностике и лечении различных заболеваний, которые на поверку могут быть, мягко говоря, совсем необоснованными, то отдельные медицинские организации должны очень вдумчиво подходить к принятию решений о внедрении каждой разработки. Что прежде всего требует оценки: вопросы, касающиеся организационного управления и технического обслуживания в период эксплуатации, а также клинические соображения и экономические обоснования. Учреждения здравоохранения должны незамедлительно начинать создавать или адаптировать уже существующие мощности и структуры автоматизированных систем управления для развертывания будущих платформ ИИ.

Руководящие органы системы здравоохранения при участии всех заинтересованных сторон в ходе проведения краткосрочного и долгосрочного планирования должны разработать регламенты, которые потребуются для измеримого улучшения производственных процессов медицинских учреждений и достижения желаемых клинических результатов. В случае четкого определения целевых показателей будущего состояния организации, ИИ, безусловно, сможет оказать положительное влияние на отечественную систему здравоохранения посредством эффективной интеграции в электронные медицинские карты *EHR electronic health record*, программы диспансеризации населения, а также вспомогательные и смежные рабочие процессы учреждений здравоохранения. Перед

внедрением ИИ медицинский менеджмент должен оценить с привлечением всех заинтересованных сторон, в частности, врачей, пациентов, широкой общественности степень прозрачности, необходимую для работы ИИ в конкретном варианте использования. Это включает определение культурного сопротивления и всего разнообразия ограничений производственных процессов, которые могут диктовать ключевые требования к интерпретируемости и применимости для успешного развертывания платформ.

Посредством управления информационными технологиями *IT Information Technology* системы здравоохранения должны установить стандартные процедуры для наблюдения и поддержания производительности приложений ИИ, автоматизируя их, чтобы обеспечить масштабируемое добавление инструментов ИИ для различных вариантов использования. Управление *IT* должно вовлекать руководство системы здравоохранения, конечных пользователей и целевые контингенты населения для создания ценностного контента для приложений ИИ. Это будет включать аналитические сообщения и материалы для определения потенциальной экономии средств и/или клинических результатов от внедрения ИИ. Разработка и внедрение ИИ должны следовать установленным передовым научным практикам внедрения и разработки программного обеспечения. Поскольку нейросетевые технологии все еще находятся на ранних стадиях развития, медицинскому менеджменту следует сохранять здоровый скептицизм в отношении рекламируемых преимуществ ИИ. Платформы, не обладающие уникальными исследовательскими или производственными ИТ возможностями, вряд ли должны находиться в приоритете. Общественность особенно приветственно относится к активным усилиям разработчиков по совершенствованию механизмов предотвращения социальной предвзятости приложений ИИ.

### **3. ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ЗДРАВООХРАНЕНИИ: НАДЕЖДА, А НЕ ШУМИХА, ОБЕЩАНИЕ, А НЕ ОПАСНОСТЬ.**

#### **Зимы ИИ и парадоксы в моделях машинного обучения**

Предоставление медицинских услуг во всем мире продолжает сталкиваться со значительными проблемами, связанными с растущей генерацией данных и знаний. Искусственный интеллект призван улучшить систему здравоохранения на фоне взрывного роста объемов электронных баз медицинских данных, значительных достижений в области аналитики данных и растущего стремления государственных органов максимально оптимизировать расходы на поддержание общественного здоровья при обеспечении справедливого и равного доступа всех слоёв населения к медицинским услугам. Инструменты ИИ потенциально могут решать многие проблемы оказания медицинской помощи, помогая отрасли здравоохранения достигнуть состояния самообучающейся системы,



учитывающей персонализированные потребности и предпочтения обслуживаемых контингентов населения. Мы уже ранее упоминали о том, что термин ИИ включает машинное обучение, обработку естественного языка, экспертные системы, оптимизацию, робототехнику, речь и зрение, а термины инструменты, платформы, модели, алгоритмы, системы и приложения ИИ используются взаимозаменяемо.

Несмотря на то, что на страницах нашей книги приводится множество впечатляющих примеров применения медицинского ИИ, мы призываем читателей взаимодействовать с ним разумно и осторожно, чтобы избежать очередной «зимы ИИ» или еще больше не усугубить неравенство в здравоохранении. Первая зима ИИ началась в 1974 г. и длилась до 1980 года. Она была вызвана провалом ожиданий в отношении перспектив ИИ после шумихи вокруг его потенциала. Считается, что всё началось с отчёта британского математика Д. Лайтхилла *Michael James Lighthill*, в котором он жестко критиковал возможности применения искусственного интеллекта для решения прикладных задач. Так, Лайтхилл заявлял, что не было сделано ни одного открытия, которое бы привело к обещанным серьезным результатам, подразумевая, что достигнутые ранее успехи были лишь в отношении игрушечных задач, а не настоящих проблем реального мира. Ещё одним серьезным заявлением того времени стал сформулированный в 1980-ых годах парадокс Х. Моравека *Moravec's paradox*. Он обратил внимание на то, что высоко когнитивные процессы ИИ гораздо менее требовательны в плане вычислительной мощности, чем самые базовые сенсомоторные операции, и именно последние наиболее сложны в разработке. Вторая «зима» длилась с 1987 по 2000 гг. Она была вызвана замещением машин с экспертными системами, одним из первых типов технологий ИИ, рабочими станциями от таких производителей как Sun Microsystems, а позже настольными компьютерами от Apple и IBM, которые к тому времени стали быстрее и имели большую вычислительную мощность. Это совпало с упадком экспертных систем в целом, чему предшествовал период колоссальных инвестиций в них. В первой половине 1980-х вокруг этих технологий возникла целая многомиллиардная индустрия, однако, уже в начале 1990-х стало понятно, что они стали слишком дорогими и сложными в использовании по сравнению с возникшими альтернативами.

Две рассмотренные выше «зимы ИИ» не являются единственными периодами пессимистических настроений. Было и несколько других небольших волн, вызванных, например, проектами по внедрению ИИ в военные технологии, результат которых не всегда оправдывал ожидания. В целом, история развития искусственного интеллекта уже много раз видела сменяющиеся периоды хайпа и последующего разочарования. Инструменты ИИ хороши лишь настолько, насколько хороши данные, используемые для их разработки и поддержки, и существует множество ограничений с текущими источниками данных. Кроме того, суще-

ствуется реальный риск усиления текущего неравенства и недоверия, если инструменты ИИ будут разрабатываться и внедряться без продуманного упреждающего планирования, грамотного управления, укрепления доверия, прозрачности, соответствующих уровней автоматизации и регулирующего надзора.

Многие известные статистические парадоксы вездесущи в рабочих процессах машинного обучения. Парадоксы – одно из чудес человеческого познания, которое трудно поддается математике и статистике. Концептуально парадокс представляет собой утверждение, которое приводит к очевидному противоречивому выводу, основанному на исходных предпосылках проблемы. Даже самые известные и хорошо задокументированные парадоксы регулярно вводят в заблуждение экспертов в предметной области, поскольку они в корне противоречат здравому смыслу. Поскольку искусственный интеллект стремится воссоздать человеческое познание, очень часто модели машинного обучения обнаруживают парадоксальные закономерности в обучающих данных и приходят к выводам, которые на первый взгляд кажутся противоречивыми. Давайте рассмотрим некоторые известные парадоксы, которые обычно встречаются в моделях машинного обучения.

Парадоксы обычно формулируются на стыке математики и философии. Пресловутый философский парадокс, известный как корабль Тесея, задается вопросом, остается ли объект, в котором были заменены все его компоненты, в основном тем же самым объектом. Во-первых, предположим, что знаменитый корабль, на котором герой древнегреческого эпоса Тесей плыл на остров Крит, чтобы убить Минотавра, хранится в гавани как музейный экспонат. С годами некоторые деревянные детали начинают портиться и поэтому заменяются новыми. Примерно через столетие все детали оказываются заменёнными. Вопрос состоит в том, является ли реставрированный корабль по-прежнему оригинальным объектом. В качестве альтернативы предположим, что каждая из удаленных частей корабля хранится в музейных запасниках, а спустя некоторое время появляются инновационные технологии, замедляющие гниение и позволяющие вновь собрать все детали вместе, чтобы воссоздать корабль. Тогда, можно ли опять-таки считать этот реконструированный корабль оригинальным? И если да, то восстановленный корабль в гавани тоже будет оригинальным?

Область математики и статистики полна известных парадоксов. Используя пару известных примеров, легендарный математик и философ Б. Рассел *Bertrand Arthur William Russell* сформулировал парадокс, который высветил противоречие в одной из самых оригинальных идей теории множеств, сформулированных одним из величайших математиков всех времен Г. Кантором *Georg Ferdinand Ludwig Philipp Cantor*. По сути, парадокс Рассела ставит вопрос о том, будет ли «список всех списков, которые не содержат самих себя», включать сам себя. Парадокс возникает в рамках теории естественных множеств при рассмотрении множества всех

множеств, которые не являются членами самих себя. Такой набор кажется членом самого себя тогда и только тогда, когда он не является членом самого себя. Отсюда парадокс. Некоторые наборы, такие как набор всех чайных чашек, не являются членами самих себя. Другие наборы, такие как набор всех не чайных чашек, являются членами самих себя. Назовём набор всех наборов, которые не являются членами самих себя **R**. Если **R** является членом самого себя, то по определению он не должен быть членом самого себя. Точно так же, если **R** не является членом самого себя, то по определению он должен быть членом самого себя.

Как и любая форма создания знаний на основе данных, модели машинного обучения не лишены когнитивных парадоксов. Напротив, когда машинное обучение пытается вывести закономерности, скрытые в наборах обучающих данных, и подтвердить свои знания в конкретной среде, они постоянно подвержены парадоксальным выводам. Вот несколько самых громких парадоксов, которые возникают в решениях для машинного обучения.

Названный в честь британского математика Эдварда Симпсона *Edward Hugh Simpson*, парадокс Симпсона описывает феномен, при котором тренд, наблюдаемый в нескольких группах данных, рассеивается при объединении данных. Реальный случай парадокса произошел в 1973 году. Уровень приема исследовался в аспирантуре Университета Беркли. Женщины подали в суд на университет из-за гендерного разрыва при поступлении. Результаты разбирательства были следующими. Когда каждая специальность рассматривалась отдельно, например, право, медицина, инженерия, то оказывалось, что женщин принимали в аспирантуру чаще, чем мужчин! Тем не менее, среднее значение убедительно демонстрировало, что мужчин принимали гораздо чаще, чем женщин. Как такое было возможно? Объяснение здесь состоит в том, что простое среднее значение не учитывает релевантность конкретной группы в общем наборе данных. В этом конкретном примере женщины подавали заявления на специальности с низким уровнем поступления, то есть в юридические и медицинские, где принималось менее 10% абитуриентов. Поэтому получалось, что доля принятых женщин была весьма мала. С другой стороны, мужчины чаще подавали заявления на специальности с высокими показателями приема, скажем, в инженерные, где показатель приема составлял более 50%. Отсюда доля принятых мужчин была очень большой.

В контексте машинного обучения многие алгоритмы неконтролируемого обучения выводят шаблоны из разных наборов обучающих данных, что приводит к противоречиям при объединении по всем направлениям. Рассмотрим ещё один парадокс, который был предложен в 1968 г. немецким математиком Дитрихом Браесом *Dietrich Braess*. На примере перегруженных транспортных сетей Браес объяснил, что, как это ни парадоксально, строительство новой дороги дорожной сети может затруднять движение, напри-

мер, время в пути каждого водителя. Аналогичным образом, закрытие одной из дорог может сократить время в пути. Рассуждения Браеса основаны на том факте, что в равновесной игре по Нэшу у водителей нет стимула менять свои маршруты. С точки зрения теории игр, индивидуум ничего не выиграет от применения новых стратегий, если другие будут придерживаться тех же стратегий. В случае с водительской стратегией – это выбранный маршрут. При парадоксе Браеса водители будут продолжать переключаться до тех пор, пока не достигнут равновесия по Нэшу, несмотря на снижение общей производительности. Таким образом, как это ни парадоксально, закрытие дорог может облегчить заторы. Парадокс Браеса очень уместен в автономных многоагентных сценариях обучения с подкреплением, в которых модели должны вознаграждать агентов на основе конкретных решений в неизвестной среде.

Ранее уже упомянутого нами Х. Моравека *Hans Moravec* можно считать одним из величайших мыслителей в области искусственного интеллекта последних нескольких десятилетий. В 1980-х он сформулировал противоречащее интуиции предположение о том, как модели ИИ приобретают знания. Парадокс Моравека утверждает, что, вопреки распространенному мнению, рассуждение высокого уровня требует меньше вычислений, чем бессознательное познание низкого уровня. Это эмпирическое наблюдение, которое противоречит представлению о том, что большие вычислительные возможности приводят к более интеллектуальным системам. Более простой способ сформулировать парадокс Моравека состоит в том, что модели искусственного интеллекта могут выполнять невероятно сложные статистические задачи и задачи вывода данных, которые невозможны для человека. Однако многие задачи, которые являются тривиальными для человека, такие как целостное восприятие объекта, требуют крайне дорогих и высокопроизводительных моделей ИИ. Специалисты по ИТ прекрасно знают, что сравнительно легко научить компьютерные программы показывать высочайшие результаты в интеллектуальных тестах или играх, и трудно или даже невозможно привить им навыки уровня годовалого ребенка, когда дело касается широты и мобильности восприятия. С точки зрения машинного обучения парадокс Моравека актуален в аспекте трансферного обучения, которое направлено на обобщение знаний по различным моделям машинного обучения. Поэтому в наиболее продвинутых приложениях ИИ люди и алгоритмы будут неплохо сочетаться друг с другом [134].

Парадокс достоверности *accuracy paradox*, непосредственно связанный с машинным обучением, утверждает, что, как ни парадоксально, достоверность не всегда является хорошим показателем для классификации результативности прогнозных моделей. Суть этого парадокса коренится в несбалансированности наборов обучающих данных. Например, в наборе данных частота случаев категории А является доминирующей и обнаруживается в 99%. При этом прогноз о том, что каждый относящийся к категории А случай



будет иметь достоверность 99%, полностью вводит нас в заблуждение. Более простой способ понять *accuracy paradox* заключается в том, чтобы найти баланс между точностью и полнотой / чувствительностью в моделях машинного обучения. В алгоритмах машинного обучения **точность** часто определяется как измерение того, какая часть наших прогнозов для положительного класса верна. Она формулируется как отношение: *истинно положительные / истинно положительные + ложно положительные результаты*. Кроме того, показатель **полноты** измеряет, как часто наши прогнозы действительно соответствуют положительному классу. Он формулируется как отношение: *истинно положительные / истинно положительные + ложно отрицательные результаты*.

Во многих моделях машинного обучения баланс между точностью и полнотой является лучшим показателем достоверности. Например, в случае алгоритма обнаружения мошенничества полнота является более важным показателем. Очевидно, что важно выявить все возможные случаи мошенничества, даже если это означает, что для этого поначалу может потребоваться пройти несколько раундов неудачных срабатываний алгоритма. С другой стороны, если алгоритм специально создан для анализа настроений пациентов и все, что нам нужно, это высокоуровневое представление об их эмоциональном состоянии, отражаемом в их высказываниях или постах в социальных сетях, тогда стремление к точности является именно той правильной стратегией, который и нужно следовать.

Напоследок коснёмся проблемы *обучаемости ИИ*, которая имеет непосредственное отношение к парадоксу Геделя. Этот парадокс связывает способность платформы ИИ к обучению с одной из самых спорных теорий математики, так называемой, теоремой Геделя **о неполноте** *Gödel's incompleteness theorem*. К. Гедель *Kurt Friedrich Gödel* – выдающийся математик, которому удалось раздвинуть границы математической логики, опубликовав 2 теоремы, в которых, по сути, говорится, что некоторые утверждения нельзя доказать, как истинные, или ложные с помощью стандартного математического языка. Другими словами, он доказал, что одного лишь языка математики явно недостаточно для понимания законов Вселенной. Эти теоремы стали известны как гипотеза континуума Геделя.

Израильские исследователи искусственного интеллекта связали гипотезу континуума Геделя с обучаемостью модели машинного обучения. В своём смелом заявлении, бросающем вызов всем общепринятым представлениям о неопределенности в обучаемости, они утверждают, что если континуум-гипотеза верна, то даже небольшой выборки будет вполне достаточно для экстраполяции. Однако, если эта гипотеза неверна, то никакой конечной выборки никогда не будет достаточно. Таким образом они показывают, что проблема обучаемости эквивалентна континуум-гипотезе, то есть первой проблеме Гильберта или проблеме Кантора о мощности континуума. Обобщённая континуум-гипотеза утверждает, что для любого бесконечного множества  $S$  не существует таких множеств,

кардинальное число которых больше, чем у  $S$ , но меньше, чем у множества всех его подмножеств  $2S$ . Следовательно, проблема обучаемости как бы находится в состоянии неопределенности, которое может быть разрешено только путем выбора аксиоматического универсума. Короче говоря, выводы израильских исследователей говорят о том, что поскольку обучаемость ИИ зависит от континуум-гипотезы Геделя, то многие проблемы ИИ могут оказаться фактически неразрешимыми. Хотя сегодня этот парадокс очень мало применим к реальным проблемам искусственного интеллекта, он будет иметь первостепенное значение для развития этой области в ближайшем будущем.

Теперь мы понимаем, что логические парадоксы вездесущи в реально стоящих перед машинным обучением задачах. Несмотря на то, что сами по себе программные алгоритмы не обладают здравым смыслом и поэтому якобы неуязвимы для статистических парадоксов, но, учитывая, что большинство проблем машинного обучения всё-таки требуют человеческого вмешательства, ибо основаны на созданных людьми наборах данных, то мы всегда будем вынуждены считаться с существующей в нашей вселенной системой парадоксов и противоречий. Специалисты компании Microsoft считают, что увеличение разнообразия представителей типов информационного метаболизма, как это было описано в монографии “*Socionic vision on Bioethics and Deontology*” (V. Berdutin, 2018), в командах разработчиков ИИ необходимо для снижения предвзятости алгоритмов. Хотя само по себе типологическое многообразие команды не способно полностью избавить ИИ от предвзятости из-за проблем, связанных с обучающими данными, тем не менее, интеграция принципа *инклюзивности и разнообразия / Diversity and Inclusion* в модели ИИ с самого начала может помочь нейросетям лучше поддерживать данные ценности.

Всё началось с того, что в начале 2023 г. Microsoft столкнулась с PR-кризисом, пытаясь продемонстрировать свои достижения в области ИИ после значительных инвестиций в OpenAI. Компания интегрировала чат-бота на базе ИИ в поисковую систему Bing, став одной из первых крупных технологических компаний, внедривших ИИ в свои основные продукты. Однако в скором времени с внедрением возникли сложности, связанные с тем, что пользователи начали массово делиться сообщениями Bing, в которых содержались неприятные высказывания и расовые оскорбления, что привлекло внимание мировой общественности. В ответ Microsoft незамедлительно ограничила функциональность ИИ, а позже заменила чат-бота Bing на Copilot. Несмотря на эти неудачи, Microsoft по-прежнему стремится использовать ИИ для обеспечения равенства и представительства при соблюдении соответствующих мер предосторожности. Компания считает, что повышение уровня разнообразия и инклюзивности в командах, разрабатывающих технологии ИИ, является ключевым фактором для устранения предвзятости в системах. Руководство Microsoft

убеждено, что для того, чтобы системы ИИ, включая большие языковые модели, такие как ChatGPT, достигали эмпатии, релевантности и точности, они должны разрабатываться людьми с разнообразными психотипами. Типологическая диверсификация команд, занимающихся разработкой ИИ, несомненно, является шагом вперед, однако и он не гарантирует абсолютного решения проблемы предвзятости, поскольку, например, такие языковые модели, как Copilot, ChatGPT и Gemini, обучаются на обширных интернет-данных, которые обязательно будут содержать те или иные представленные в общественном сознании предрассудки и предвзятости, которые невозможно искоренить никакими способами [123, 125].

Итак, мы ознакомились с некоторыми современными научными воззрениями, которые могут оказаться полезными при написании информационно-справочной и сопроводительной документации платформ медицинского ИИ и к тому же выглядят интересными для всех категорий стейкхолдеров: разработчиков, медицинских работников, пациентов, их представителей, научных экспертов, бизнес-аналитиков, а также общественных организаций и регулирующих органов. Таковая документация необходима, чтобы:

1. определиться с выбором текущих или перспективных вариантов приложений ИИ для развёртывания в учреждении здравоохранения;
2. узнать о ключевых проблемах, ограничениях и передовом опыте разработки, внедрения и обслуживания платформ ИИ;
3. ознакомиться с актуальной нормативно-правовой базой по теме ИИ;
4. обеспечить равенство, инклюзивность, соблюдение прав пациентов и медицинского персонала при взаимодействии с ИИ;
5. установить для себя приоритеты в области ИИ.

В следующем разделе мы постараемся предоставить читателям обобщенную картину проблем и достижений в области нейросетевых технологий, выделив текущие передовые практики и ключевые приоритеты в этой сфере.

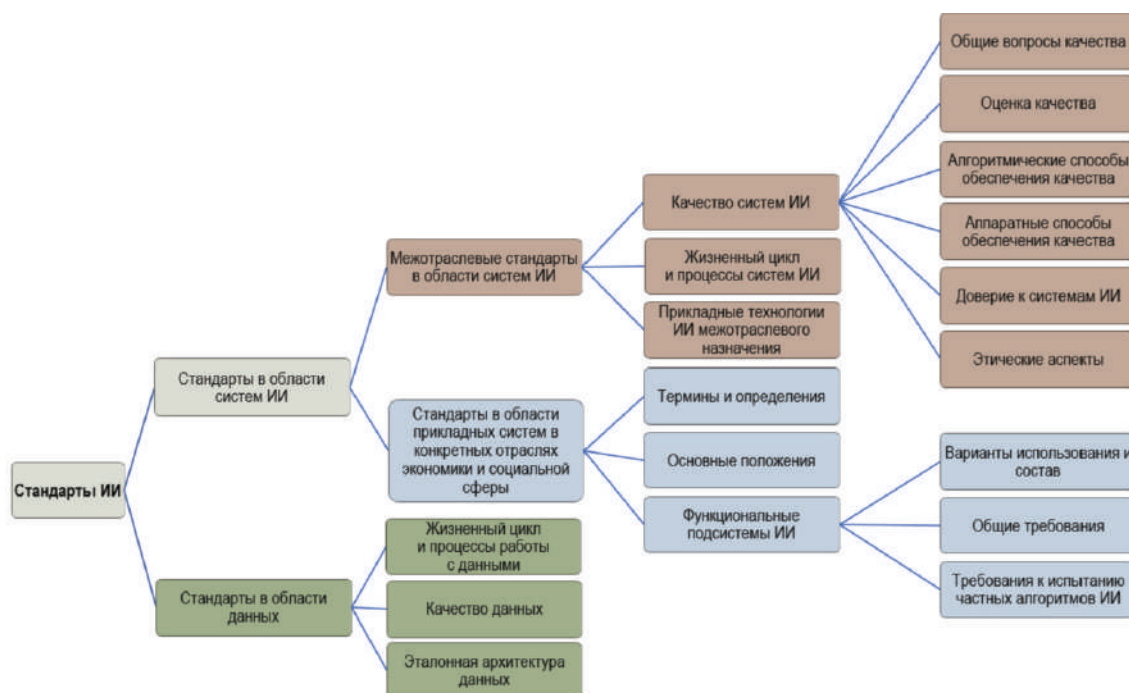
### **Проблемы и ключевые приоритеты в области стандартизации ИИ.**

Итак, мы с вами выяснили, что для успешной разработки систем ИИ требуются высококачественные, репрезентативные и разнообразные данные, которые должны подаваться платформе ИИ в стандартизованном виде. На рис. 6 показана структура системы национальных стандартов РФ в области искусственного интеллекта.

Деятельность по стандартизации искусственного интеллекта в Российской Федерации ведется созданным в 2019 г. техническим комитетом по стандартизации «Искусственный интеллект» (ТК 164), позиционирующимся как зеркальное отражение на национальном уровне профильного международного подкомитета ISO/IEC JTC 1 SC 42 Artificial Intelligence и объединившим более

100 заинтересованных государственных, коммерческих, научных и образовательных организаций и разработчиков в области технологий искусственного интеллекта и больших данных. Стандартизация ИИ в России началась с существенным отставанием от международных организаций и зарубежных стран-лидеров в этой области, в связи с чем на национальном уровне было принято решение ускоренными темпами сократить разрыв, разработать и ввести в действие ряд основополагающих национальных стандартов в области больших данных, гармонизированных с международными. Основные направления стандартизации ИИ в РФ после тщательного анализа международного контекста были зафиксированы в Перспективной программе стандартизации в области приоритетного направления «Искусственный интеллект» на 2021-2024 гг. утвержденной в 2020 г. Минэкономразвития России и Росстандартом. Программа реализуется в рамках федерального проекта «Искусственный интеллект» и предполагает разработку в течение 4-ех лет более 200 стандартов, обеспечивающих опережающее развитие данного высокотехнологичного направления [20, 21, 90].

Структура Программы отражает представления ТК 164 о направлениях стандартизации ИИ, учитывающие необходимость стандартизации процессов разработки и использования технологий ИИ, при этом не вступая в противоречие с действующими нормативными техническими документами и не дублируя их. Особенностью российской Программы является не только выделение группы стандартов межотраслевого применения (универсальные стандарты управления жизненным циклом и качеством



**Рис. 6.** Структура системы национальных стандартов РФ в области искусственного интеллекта.



систем ИИ, стандарты систем ИИ межотраслевого назначения), но и выделение в отдельную группу стандартов, в которых так или иначе учитывается специфика решаемой прикладной задачи (стандарты для прикладных систем ИИ в конкретных отраслях экономики и социальной сферы). Наконец, третью группу образуют стандарты в области больших данных, используемых при создании и применении систем ИИ (Гарбук С.В. с соавт., 2021) [13].

Вошедшие в Программу «межотраслевые» стандарты в большинстве своем представляют гармонизацию стандартов, уже принятых или разрабатываемых на международном уровне, прежде всего в ПК42. Особое внимание при этом уделяется универсальным подходам к оценкам качества (и его составляющих – результативности, эффективности, безопасности) систем ИИ. При этом фиксируются требования и технические спецификации как к используемым вычислительным средствам, так и к применяемым алгоритмам и моделям. Учитывая значение данных в успешном функционировании систем ИИ, большое внимание в России уделяется стандартизации работы с большими данными, включая качество данных для аналитики и машинного обучения, выделенной в отдельное направление Программы.

Особенностью российского подхода также стало особое внимание стандартизации прикладных систем ИИ в конкретных сферах деятельности, в частности, в здравоохранении, для которого формируется своя группа стандартов, построенная по единому принципу: онтология использования ИИ; описания типовых информационно-технологических систем, решающих характерные для медицины задачи и общие вопросы создания и использования функциональных подсистем, использующих технологии ИИ. Применительно к последним объектам стандартизации формируются требования к интеллектуальным алгоритмам и моделям, эталонным архитектурам, интерфейсам взаимодействия со смежными подсистемами, инфраструктуре разработки и развертывания подсистем. Неотъемлемой частью отмеченной группы стандартов являются документы, содержащие требования к испытаниям, показатели и критерии качества функциональных подсистем, варианты их использования и требования к наборам тестовых данных, включая сами наборы и/или процедуры их корректного расширения. По состоянию на середину 2023 г. в России утверждено более 50 национальных стандартов.

Основополагающие стандарты представлены у всех международных организаций стандартизации и представляют собой базу для последующих технических спецификаций и формирования единого терминологического аппарата, позволяющего практикам говорить на одном языке между собой, с разработчиками или регуляторами. Наряду со сквозными / горизонтальными стандартами, фиксирующими сложившиеся в данный момент фундаментальные концепции, устоявшиеся определения и фундаментальные представления об ИИ, используемые в различных сферах деятельности, имеется вертикальное направление, связанное со стандартизацией, учитываю-

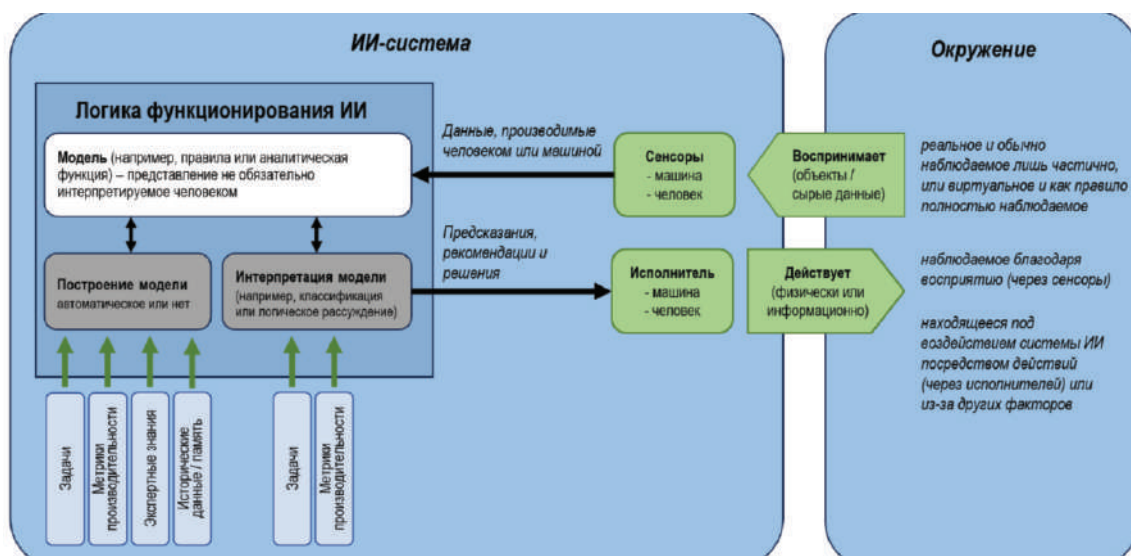
шей специфику конкретной сферы деятельности или предметной области ИИ. Вертикальные спецификации сосредоточены на решении задач, относящихся только к данному применению технологий ИИ или к отдельной сфере деятельности (Хохлов Ю. Е., 2023) [37].

Несмотря на то, что формально Организация экономического сотрудничества и развития/ОЭСР не занимается стандартизацией, результаты исследований и методологические подходы, разработанные её экспертами, учитываются международными организациями по стандартизации. Примером может служить рекомендация Совета по искусственному интеллекту сопоставление с которой нашло отражение в основополагающем стандарте ИСО/МЭК «Искусственный интеллект. Концепции и терминология искусственного интеллекта» (OECD, 2022). Разработанное Экспертной группой по искусственному интеллекту ОЭСР высокоуровневое описание произвольной системы искусственного интеллекта приведено на рис. 7:

Несмотря на огромное количество наработок и материалов по стандартизации ИИ некоторых инструментов оценки качества и методологии курирования данных пока не хватает. Функциональная совместимость имеет решающее значение на всех уровнях, в том числе: в электронных медицинских картах, вспомогательных компонентах системы здравоохранения, между медицинскими учреждениями и различными потребительскими приложениями. Мы не можем игнорировать тот факт, что существуют различные требования к данным для обучения ИИ и для последующего использования ИИ.

Методы оценки достоверности и воспроизводимости данных часто носят случайный характер. В конечном счете, для того чтобы модели ИИ были достоверными, семантика и происхождение данных, используемых для их получения, должны быть полностью прозрачными, недвусмысленно переданными и доступными, по крайней мере, для проверки независимым проверяющим агентом. Это отдельный элемент прозрачности, и смешение прозрачности данных с алгоритмической прозрачностью усложняет дискурс экосистемы ИИ. Мы обращаем внимание читателей на четкое разделение этих тем. Одним из примеров декларации принципов, способствующей надежности и качеству данных, являются принципы FAIR *findability, accessibility, interoperability, reusability* / находимость, доступность, совместимость и возможность повторного использования. Эти принципы, выдвинутые исследователями в области молекулярной биологии и биоинформатики, нелегко формализовать или реализовать. Однако для того, чтобы медицинский ИИ стал по-настоящему зрелым продуктовым направлением ИТ, необходимо строго придерживаться этого набора принципов.

Медицинскому сообществу следует активнее содействовать разработке законодательных механизмов, облегчающих агрегацию данных, включая внутриотраслевой идентификатор медицинских услуг и стандартные механизмы объединения данных из ответственных источников. Дебаты должны быть сосредоточены на глубоко продуманной концепции о способности крупномасштабных информационных



**Рис. 7.** Концептуальное описание системы ИИ.

ресурсов здравоохранения служить общественному благу и далеко идущих последствиях применения этой концепции. Дискуссии о доступности и репрезентативности медицинской информации должны быть тщательно сбалансированы научным подходом к пропаганде и образованию с целью достижения общественного консенсуса по вопросам: 1) где и как персональные данные могут быть использованы для нужд ИИ, 2) монетизация данных, 3) какие стейкхолдеры могут повторно использовать персональные данные и в каком объеме, 4) какие гарантии безопасности должны быть приняты.

Когда разработчики ИИ пытаются объединять данные из разных источников, они сталкиваются с техническими, законодательными и этическими препятствиями. Существует настоятельная потребность в смелых инновационных решениях, которые позволяли бы надежно агрегировать данные, не теряя их прозрачности и одновременно уважая конфиденциальность и чувства пациентов. Кроме того, эти решения должны быть справедливыми, чтобы даже наименее подкованные в компьютерных технологиях пациенты хорошо понимали, что их данные приносят пользу всему обществу. Известный статистик и основатель математической статистики К. Пирсон говорил: «То, что измеряется, улучшается». Поэтому приоритет справедливости и инклюзивности должен стоять во главе угла при развертывании платформ медицинского ИИ. Разработчикам крайне важно учитывать слабости и недостатки обучающих данных, чтобы заранее исключать проявления предубеждений ИИ. Также важно учитывать, как следует внедрять приложение ИИ, и может ли среда развертывания повлиять на справедливость и инклюзивность, поскольку практически повсеместно мы сталкиваемся с неравенством в доступности к медицинской помощи из-за существующих социальных детерминант, искажающих базовые принципы систем здравоохранения многих стран. К несчастью, в обществе потребления любая технологическая новинка часто ус-

губляет исторически сложившееся неравенство в социальной сфере. Цифровое неравенство продолжает оставаться реальностью и в эпоху носимых медицинских устройств, даже если первоначально их производство субсидируется. ИИ может усугубить неравенство в колоссальных масштабах, поскольку влияние одного предвзятого человека гораздо меньше, чем влияние глобального или национального ИИ. Например, Google в погоне за имиджем запретил своим программистам выявлять предвзятость ИИ поисковой системы, которая отслеживалась путём создания множества фейковых профилей (O’Neil, 2016) [126].

Транспарентность данных является ключом к тому, чтобы независимые эксперты могли изучить исходную информацию на предмет предвзятости и убедиться в её репрезентативности. К сожалению, во многих случаях ИИ имеет дело с данными, которые не являются репрезентативными для обслуживаемого населения. Порой всё же наблюдаются робкие попытки связать их с социальными детерминантами здоровья, окружающей среды и данными социальных сетей для получения всеобъемлющего профиля пациентов. Но эти попытки осуществляются вне этического или правового контекста. Крайне важно, чтобы подходы к оценке отчетности по качеству и репрезентативности данных были стандартизированы. Наиболее перспективным выглядит использование характеристик информационного метаболизма членов команд разработчиков ИИ, чтобы извлекать максимальную выгоду из их столь разнообразных мировоззрений, знаний и способов мышления (Бердугин В.А., 2020). Обеспечение справедливости и инклюзивности потребуют новой управленческой парадигмы. Предпринятые усилия в этом направлении со стороны технологических ИТ-гигантов сопровождались многочисленными скандалами, трудностями и неудачами. Показательным примером служит роспуск Совета по этике Google в апреле 2019 г. Не даром М. Latonero предположил, что для того, чтобы ИИ был хоть как-то общественно полезен, результаты его использования не должны наносить ущерб фундаментальным человеческим ценностям (Latonero, 2018). Возможно, лучший способ гарантировать, что справедливость и инклюзивность являются основополагающими компонентами процветающей системы здравоохранения – это добавить их в перечень целевых характеристик качественного медицинского обслуживания, поместив их в ряд с: улучшением показателей здоровья населения, передовыми практиками лечения больных, благополучием медицинского персонала и снижением не добавляющих ценность медицинских услуг расходов на их оказание [8, 100].

Ключевой проблемой принятия и широкого использования медицинского ИИ являются недостатки исходных данных: недостаточная алгоритмическая прозрачность и точность, потенциальный риск и неопределенность гражданской ответственности при их использовании. Одним из декларируемых в нашей книге приоритетов является оценка транспарентности ИИ в контексте с угрозами



его использования. Поэтому обеспечение надлежащей прозрачности в каждой подобласти данных и системный мониторинг производительности алгоритмов также должны быть приоритетами. Кроме того, медицинский менеджмент должен следить за темпами окупаемости инвестиций, прочими выгодами от внедрения и текущими рисками, включая риски неблагоприятных событий в отдаленном будущем. Инженеры-информатики, обслуживающие платформы ИИ, должны глубже погружаться в культуру и производственные процессы медицинских учреждений, корректируя алгоритм согласно запросам пользователей. Все стейкхолдеры должны вносить посильный вклад в обеспечение справедливости, инклюзивности, обновляемости и подконтрольности приложения ИИ. Названные атрибутивные характеристики отнюдь не изолированы друг от друга, они вместе составляют единый взаимосвязанный функциональный комплекс. Во всех случаях следует поддерживать прозрачность базовых данных, используемых для создания модели ИИ. Хотя персональные данные пациента не должны быть общедоступными, должна быть обычной практикой публикация информации: об источниках данных, из которых они были агрегированы; как данные были преобразованы; какие были проблемы с качеством данных; критерии включения в когорты и исключения из них; сводная статистика демографических данных. Эта вспомогательная информация значительно улучшала бы текущее понимание и доверие к инструментам ИИ.

Необходимость в алгоритмической прозрачности во многом зависит от контекста использования. Для приложений, которые непосредственно влияют на качество жизни пациентов или результаты лечения, базовые требования к прозрачности высоки. Однако уровень прозрачности может различаться в зависимости от 1) известной точности ИИ; 2) ясности рекомендуемых действий для конечных пользователей; 3) риска для пациента или цели исследования; 4) возникновения юридической ответственности. Например, если инструмент ИИ имеет высокую точность и низкий риск, предоставляет четкие рекомендации конечному пользователю и вряд ли влечёт юридическую ответственность кого-либо из стейкхолдеров: учреждения здравоохранения, производителя программного обеспечения или конечного пользователя, то потребность в полной алгоритмической прозрачности, вероятно, будет ниже.

### **Конкуренция нейросетей в области дифференциальной диагностики опухолей мозга.**

Опухоли мозга является широко распространенной патологией, которая приводит к многочисленным летальным исходам по всему миру, включая развитые страны Запада; только в Соединенных Штатах зафиксировано около 20 тыс. смертей. Новообразования мозга могут иметь как доброкачественную, так и злокачественную природу, иметь разный прогноз и скорость течения, поэтому диф-

дифференциальная диагностика вида опухоли является крайне актуальной задачей для системы здравоохранения. Менингиомы, составляющие 36,1% всех первичных опухолей, имеют тенденцию группироваться в виде бугорков на внешних контурах головного и спинного мозга, поражая покрывающие его оболочки. Менингиомы можно диагностировать на основе их анатомического расположения, формы и вида клеток. Несмотря на медленное развитие, эти опухоли могут вызывать серьезные побочные эффекты, включая судороги и потерю зрения. Опухоли, развивающиеся не из нейронов, а из глиальных клеток мозга называются глиомы. Они включают в себя большую группу опухолей с широким спектром симптомов и уровней опасности. Гипофизарные опухоли, развиваются из клеток гипофиза, расположенного у основания черепа и играющего жизненно важную роль в гормональной регуляции. Они могут формировать там аберрантные клеточные массы, которые могут крайне негативно влиять на многие биологические системы организма.

Способность успешно диагностировать и лечить опухоли мозга зависит от полного понимания всех аспектов развития патологии. Радиологи используют различные методы визуализации, такие как ультразвук, электроэнцефалография, рентгенография, компьютерная томография/КТ, магнитно-резонансная томография/МРТ, ПЭТ-КТ для исследования опухолей мозга, помощи в точной диагностике и выборе соответствующего лечения. Первичные опухоли мозга значительно различаются по размеру, местоположению и другим признакам, что затрудняет раннюю диагностику. Качество визуализации опухолей неинвазивными методами зависит от абсорбционных характеристик тканей, что делает необходимым точное определение скоростей поглощения излучения. При сравнении различных вариантов визуализации МРТ выделяется как золотой стандарт, поскольку она может предоставить ценную информацию о мозге как в здоровом, так и в больном состоянии. В случае наличия аномалий она помогает точно определить вид опухоли. Однако анализ результатов МРТ-сканирования требует огромного опыта и высокого уровня профессионального мастерства, что делает затруднительным его качественное выполнение рядовым врачом. К сожалению, многие отечественные медицинские учреждения по-прежнему не обладают специалистами достаточно высокого экспертного уровня знаний для осуществления точной дифференциальной диагностики опухолей мозга.

Своевременное обнаружение опухолей мозга является сложной задачей, но развитие алгоритмов глубокого обучения показало многообещающие результаты в точном обнаружении опухолей мозга с использованием цифровых изображений. МРТ и КТ в сочетании с моделями глубокого обучения показали высокую эффективность в деле обнаружения и диагностики опухолей мозга. Модели глубокого обучения, основанные на сверточных нейронных сетях, рекуррентных нейронных сетях, автокодировщиках, трансферном обучении и

гибридных алгоритмах, активно изучаются исследователями в последние годы из-за их способности к высокой точности и раннему выявлению патологии. В частности, сверточные нейронные сети *Convolutional Neural Networks / CNN* стали у онкологов методом выбора для обнаружения и классификации опухолей мозга. Эти передовые модели успешно решают сложные проблемы компьютерной диагностики *Computer-Aided Diagnosis / CAD*, охватывающие распознавание, классификацию, сегментацию и даже обнаружение. Однако многие существующие CAD-решения для обнаружения и идентификации опухолей мозга, основанные на *CNN*, страдают от недостаточной эффективности и требуют значительных вычислительных ресурсов. С одной стороны, несложные архитектурные варианты моделей *CNN* имеют непреодолимые ограничения по точному определению местоположения опухоли. С другой стороны, продвинутые и способные к сегментации сверточные модели нейросетей могут точно локализовать пораженную область с помощью маски, но лишь при условии наличия больших вычислительных мощностей. Это создает серьезные трудности при отсутствии у медицинских учреждений компьютерного оборудования экстракласса.

Сверточные нейросети обычно кодируют входное изображение в последовательность промежуточных признаков меньшей размерности. Такой подход хорошо работает для задачи классификации изображений, но плохо работает для задачи распознавания и локализации объектов. Для обхода ограничения сверточных сетей по локализации и сегментации объектов применяются *CNN* кодировщик-декодировщик с архитектурой типа «песочные часы» *Convolutional Encoder-Decoder Neural Network*. В таких архитектурах декодировщик располагается поверх кодировщика. Таким образом, кодировщик применяется для задач классификации, а декодировщик — для задач локализации и сегментации. При этом кодировщик являясь основной моделью *backbone model*, как правило, содержит больше параметров и потребляет больше вычислительных мощностей, чем декодировщик. При этом архитектура типа «песочные часы» неэффективна для генерации признаков *разных масштабов*, потому что в такой модели масштаб изображения постоянно уменьшается при помощи кодировщика.

Существует несколько алгоритмов обнаружения различных объектов на изображениях. Наиболее известные из них — *Single Shot Multibox Detector (SSD)*, *R-CNN* и *Fast R-CNN*. Для повышения производительности платформ ИИ за счет использования агрегированной информации, связанной с семантикой и положением опухоли, также могут использоваться самоконтролируемые модели глубокого обучения. Алгоритм *You-Only-Look-Once / YOLO*, использующий единую унифицированную нейронную сеть, произвел настоящую революцию в обнаружении опухолей мозга благодаря своей уникальной системе идентификации объектов. *YOLO* решает поставленную задачу методом регрессии, оценивая координаты ограничивающего опухоль прямоугольника и вероятности классов непосредственно из

информации на уровне пикселей. Алгоритм позволяет одновременно прогнозировать несколько ограничивающих прямоугольников и вероятностей классов, имея более высокую производительность с точки зрения как скорости, так и точности. Благодаря своей способности точно классифицировать и определять местонахождение аномалий на изображениях он нашел широкое применение в области дифференциальной диагностики опухолей мозга. На практике используются 2 варианта моделей глубокого обучения YOLO, а именно YOLOv5 и YOLOv7, которые обеспечивают более эффективное решение с улучшенной производительностью для сегментации и классификации опухолей головного мозга по данным МРТ.

Датасет сквозной CNN модели BrainMRNet содержит 253 МРТ изображения: 98 нормальных изображений и 155 снимков с опухолями. Оценивая модель с точки зрения стандартных параметров, можно констатировать, что она достигает точности 96%, достоверности 92% и чувствительности 96%. В качестве модели классификации 3D CNN проверяет извлекаемые признаки с помощью нейронной сети прямого распространения. Процесс извлечения признаков выполняется при помощи предварительно обученной сети VGG19. Модель обучается и оценивается с использованием наборов данных BraTS 2015, 2016 и 2018 с точностью 98%, 96% и 92% соответственно. В некоторых случаях платформа достигает и более высокой точности, но результат всегда варьируется в зависимости от накапливаемых данных.

Анализирующая МРТ изображения и основанная на технологии заплаток Dice нейросеть / DNN широко применяется для классификации опухолей мозга. Работа модели основана на восьми датасетах, включая BraTS 2012-2015, Ischemic Stroke Lesion Segmentation / ISLES 2015 и 2017 и MICCAI. DNN достигла высокого Dice коэффициента сходства 99,8% при обучении на датасете BraTS 2013, продемонстрировав постоянную производительность при различных модальностях и наборах данных. Используя метод сегментации, в котором локальные и глобальные признаки используются для диагностики, DNN в автоматическом режиме проводит обнаружение опухолей мозга. Набирая 85% очков по системе Dice, модель обеспечивает 93% специфичности и 80% чувствительности.

Для обнаружения опухолей мозга типа глиомы используется модель глубокого обучения, основанная на гибридной сверточной нейронной сети, состоящей из 2-ух и 3-ех путевых сетей. Эта модель тестировалась на датасете BraTS 2013, достигнув уровня показателей: 86% очков по системе Dice, 86% чувствительность и 91% специфичность. Набор данных содержал сканы 20 пациентов с четырьмя модальностями, что составило в общей сложности 80 изображений. Есть положительный опыт применения системы поддержки принятия решений для классификации опухолей мозга нескольких классов, которая основана на предварительно обученной модели глубокого обучения Densenet201. Для выбора признаков там используются 2 различных метаэвристических алгоритма.



Модель обучается и оценивается с использованием датасетов BraTS 2018, BraTS 2019 и классификатора SVM, достигая высокой точности 95%. Особенностью работы платформы является настраиваемое двойное подавление кодирования и факторизованное билинейное кодирование, подключенное к стандартной CNN модели. Аналогичным образом, работает основанная на трансферном обучении остаточная нейросетевая платформа, которая была проверена с точки зрения точности с помощью стандартных моделей, включая ConvNet, AlexNet и VGG 16. Она также достигла точности 95% на датасете BraTS.

В инновационном подходе глубокого обучения, включающем методы трансферного обучения, был представлен впечатляющий уровень точности 99,68% с использованием ResNet50-v2. При этом разработана автоматизированная система идентификации и сегментации опухолей, достигающая точности сегментации 95% за счет использования изображений масок в качестве меток. Платформа CNN выполняет предварительную обработку, сегментацию и классификацию, демонстрируя высочайшую эффективность по сравнению с альтернативными методами. Однако всё же стоит отметить, что и здесь, к сожалению, иногда наблюдаются некоторые проблемы при проведении сегментирования. Встроенный в архитектуру CNN и имеющий среднюю точность 77,60% алгоритм обнаружения опухолей Faster R-CNN при их классификации может достигать весьма высокой точности 96,56%. Эффект достигается путём объединения предварительно обученной CNN с серыми характеристиками матрицы совпадений. Ограниченность имеющихся сведений, касающихся данной платформы ИИ, препятствует всестороннему анализу её общей производительности. В отсутствии таких важных метрик, как отклик и конкретные пороги обнаружения, довольно сложно правильно оценить и эффективность связанной работы обеих моделей.

Платформа обнаружения объектов на медицинских изображениях YOLO заслуживает самых высоких оценок со стороны пользователей благодаря своей высокой скорости и производительности. Версия YOLOv5, интегрированная в портативную систему визуализации головы с помощью микроволн, автономно классифицирует и обнаруживает аномалии человеческого мозга. Используя 400 образцов изображений, охватывающих неопухолевые и опухолевые случаи различных локализаций, YOLOv5l выдает показатели: точность 96,32%, достоверность 95,17%, чувствительность 94,98%, специфичность 95,28%, F1-оценка 95,53% и, наконец, средняя точность *Mean Average Precision / mAP*, которая дает полное представление о производительности модели на разных уровнях сложности обнаружения – 96,12%. При переносе обучения YOLOv5 на датасет BraTS 2020 платформа достигла точности 82-92% в зависимости от версии YOLO, причем YOLOv5 показала самый высокий уровень точности. Здесь был достигнут компромисс между максимально воз-

можной точностью и временем обучения модели. Состоящая из 225 слоев и 7 413 608 параметров платформа YOLOv5 представляет собой надежную архитектуру глубокой сверточной нейронной сети, в которой различные слои извлекают значимые признаки, а локализация объектов достигается в изображениях, как показано на рис. 8.

Conv – свёртывание;

C3 – CSPNet / Cross Stage Partial Network работает на фреймворке Darknet. Метод применяется не сам по себе, а как улучшение уже существующих остаточных нейросетей *residual neural networks* / ResNet. Основная концепция заключается в том, чтобы поток градиента распространялся по разным сетевым путям за счет разделения потока градиента. Таким образом распространяемая информация о градиенте может иметь большую корреляцию, если переключать этапы конкатенации и перехода. CSPNet может значительно сократить объем вычислений, повысить скорость вывода и точность. Суть CSPNet заключается в более сложной обработке пирамид признаков;

Concat – каскадное соединение;

SPPF – *spatial pyramid pooling fast*, пространственная пирамида, состоящая из обобщающих слоёв.

Представленная выше модель начинается с серии сверточных Conv слоев, которые изучают низкоуровневые признаки из входных изображений. За этими слоями следуют слои C3, которые помогают в агрегации признаков и слиянии информации. Слои C3, отвечающие за сбор более общей и высокоуровневой информации, используют более крупные фильтры. Понижение частоты дискретизации достигается с помощью слоев Conv с настройками шага и размера ядра для улучшения способности модели собирать многомасштабные признаки путем снижения пространственных измерений карт признаков. Входное изображение понижается, чтобы модель могла

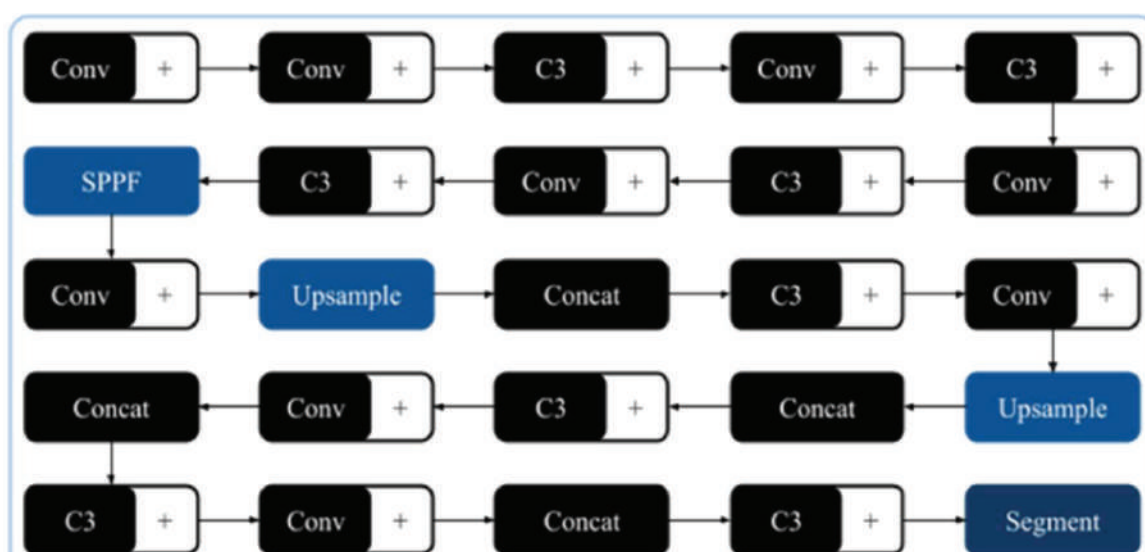
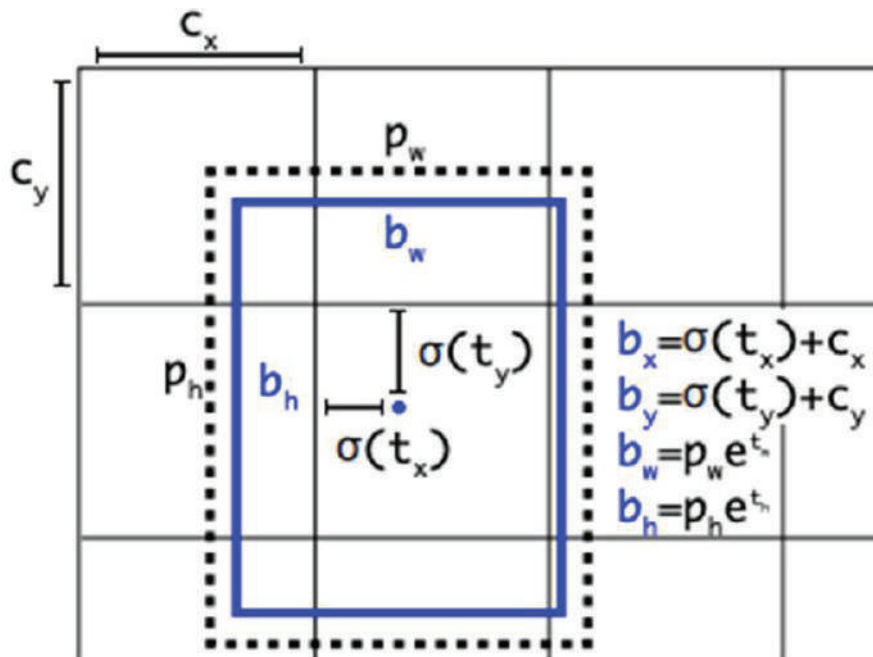


Рис. 8. Архитектура YOLOv5 для сегментации и классификации опухолей головного мозга.

лучше идентифицировать элементы разных размеров. Компонент *Spatial Pyramid Pooling Fusion / SPPF* также является частью дизайна YOLOv5. Собирая контекстную информацию в различных масштабах, он выполняет операции объединения различных размеров на картах объектов. Это улучшает понимание алгоритмом содержимого изображения, что повышает точность распознавания объектов. Последние уровни модели YOLOv5 – это слои *Conversion / Conv* и *Concatenation / Concat*. Слои Conv помогают в тонкой настройке признаков, в то время как слои Concat объединяют разномасштабные карты признаков для создания подробных представлений объектов. Последний слой модели YOLOv5, называемый YOLOv5 Segment, выполняет фактическую задачу обнаружения объектов.

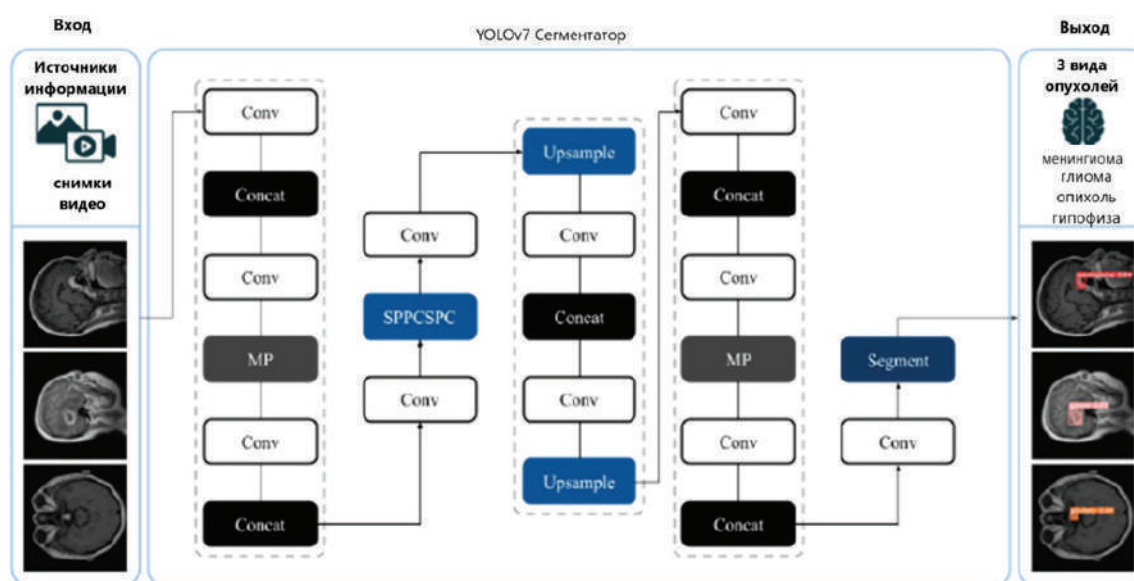
Он генерирует прогнозы ограничивающих рамок для трех масштабов, позволяя модели обнаруживать объекты различных размеров и соотношений сторон. Эти ограничивающие рамки прогнозируются на основе опорных рамок. Математические формулы, используемые для прогнозирования ограничивающих рамок в YOLO, включают определение ширины и высоты как смещений от центроидов кластера и координат центра относительно местоположения применения фильтра с использованием сигмоидальной функции. В частности, сеть предсказывает пять координат для каждого ограничивающего прямоугольника:  $t_x$ ,  $t_y$ ,  $t_w$ ,  $t_h$  и  $t_0$ . Если ячейка расположена со смещением от верхнего левого угла изображения на  $c_x$ ,  $c_y$ , а ограничивающие априорные значения имеют ширину и высоту, представленные как  $p_w$ ,  $p_h$ , то прогнозы соответствуют  $b_x$ ,  $b_y$ ,  $b_w$ ,  $b_h$ , как показано на рис. 9.



**Рис. 9.** Ограничивающая рамка с априорным измерением и прогнозированием местоположения.

YOLOv7 представляет собой ещё более усовершенствованный вариант фреймворка YOLO, обеспечивающий достижение высокоэффективных и точных результатов. Он включает в себя ряд инновационных решений для повышения производительности и эффективной обработки объектов в разных масштабах. Он отлично справляется с захватом контекстной информации из разных масштабов за счет интеграции многомасштабного слияния признаков, что повышает точность обнаружения объектов. Кроме того, пространственное пирамидальное объединение позволяет модели извлекать информацию в разных масштабах и эффективно обрабатывать объекты разных размеров. Более того, он интегрирует слои повышения дискретизации для облегчения обнаружения более мелких объектов за счет увеличения карт признаков. Он использует операции конкатенации и несколько сверточных слоев для итеративного улучшения представления объектов и повышения точности прогнозов ограничивающих рамок. На рис. 10 показана используемая архитектура сегментатора YOLOv7 для сегментации и классификации опухолей головного мозга. Сегментатор состоит из 82 слоев свертки, 15 слоев конкатенации, 5-ти слоев максимального пула, слоя *SPPCSP spatial pyramid pooling-Cross Stage Partial*, двух слоев повышения дискретизации и слоя сегмента в качестве основы алгоритма сегментации опухолей мозга.

Платформа ИИ прошла обучение на 2144 изображениях с весами *yolov7-seg.pt* для 100 эпох. Нейросеть точно дифференцирует менингиому, глиому и опухоли гипофиза на основе стационарных МРТ изображений и видеоконтента. Работа алгоритма начинается с инициализации оптимизированных весов сегментатора, определения размеров объекта и оценки достоверности. Затем каждый слой выполняет свойственные для него функции: сверточный слой проводит



**Рис. 10.** Архитектура YOLOv7 для сегментации и классификации опухолей головного мозга.



извлечение признаков путем фильтрационной свертки входных данных, слой конкатенации объединяет карты признаков с различных этапов работы нейросети, слой максимального пулинга *Max Pooling / MP* разделяет входные данные на неперекрывающиеся области и сохраняет максимальное значение в каждой области, слой SPPCSPC эффективно захватывает многомасштабную информацию с помощью слоя SPP и гарантирует надлежащий поток градиентов и информации через различные этапы сети с помощью соединения CSP / Cross Stage Partial Network, слой повышения частоты дискретизации улучшает пространственное разрешение карт признаков, а слой сегмента генерирует прогнозы для границ сегментации и вероятностей классов для сегментированных объектов. После обработки всех слоев на выходе получается сегментированное изображение опухоли головного мозга (Almufareh, et al., 2024) [45].

В таблице 3 представлены самые значимые из существующих инструментов, технологий и моделей ИИ, используемых для обнаружения и дифференциальной диагностики опухолей мозга.

Итак, мы рассмотрели использование различных нейросетевых платформ ИИ, используемых для обнаружения, классификации и дифференциальной диагностики опухолей головного мозга. Точность этих моделей зависит от таких факторов, как сложность базового алгоритма, качество обучающих датасетов и природа извлекаемых признаков. Среди рассмотренных здесь платформ модели класса YOLO имеют наилучшую производительность при решении повседневных задач. В отношении несложных классификационных моделей CNN следует отметить ограниченность их применимости при необходимости точного определения природы мелких патологических образований и объектов. Модели сегментации, такие как YOLO, устраняют это ограничение, хотя, к сожалению, имеют более высокую стоимость обслуживания. Целью данного раздела является предоставление читателям сбалансированного взгляда на проблему оптимального выбора платформ ИИ, функционирующих на базе интеграции новейших алгоритмов обнаружения патологических новообразований с алгоритмами их классификации.

### **Ближайшие перспективы медицинского ИИ**

Хотя некоторые приложения ИИ для бизнес-операций в сфере здравоохранения, вероятно, в самое ближайшее время будут полностью готовы к работе в автономном режиме, фокус внимания медицинской общественности должен быть сосредоточен на разработке, оценке и продвижении тех инструментов ИИ, которые поддерживают наши познание и деятельность, а не подменяют их. Массовая культура и маркетинг перегрузили термин *искусственный интеллект / artificial intelligence AI* до такой степени, что он у многих людей стал ассоциироваться с массовой заменой людей на роботов [129]. В результате для обозначения ИИ, который используется именно для поддержки человеческого познания, появились новые термины. К

Таблица 3.

**Инструменты, технологии и модели ИИ, используемые для обнаружения  
и дифференциальной диагностики опухолей мозга**

Модель	Датасет	Функция	Производительность	Примечание
3D CNN	BraTS 2015(192 Train, 82 Test),  BraTS 2017(285 Train, 146 Test),  BraTS 2018(285 Train, 191 Test)	классификация	98% Точность 2015  96% Точность 2017  92% Точность 2018	Вариативность показателей зависит от соотношения обучающего датасета и результатов тестирования модели
BrainMRNet сверточная нейросеть	253 изображения, 2 класса, датасет, содержащий MPT изображения экспертного уровня	классификация	96% Точность  92% Достоверность 96% Отклик/полнота	Превосходит стандартные модели, но имеет маленький датасет
Патчируемая Dice нейросеть глубокого обучения	BraTS2015 (274 обучающих и 110 тестовых циклов)	сегментация	95,1% Точность  97,2% Достоверность  95,2% Отклик/полнота	Постоянная результативность при различных датасетах и модальностях
Hybrid CNN  Гибридная сверточная нейросеть	BraTS2013	сегментация	85% очков Dice  86% отклик  91% специфичности	Избыточная сложность из-за большого количества этапов обучения
Билинейная нейросеть с индивидуальным шифрованием файлов	Датасет ВТ с двумя различными наборами данных	классификация	Результаты моделирования не содержат информации о точности, достоверности и отклике/полноте, особенно при рассмотрении всех классов в совокупности.	Любопытная архитектура, но неустойчивая производительность при разных датасетах
Глубокая остаточная нейросеть с трансферным обучением	Датасет ВТ	классификация	95% Точность  94% Максимальная достоверность  95% Полнота	Более низкая точность по сравнению с ResNet18 и ResNet101 при большой сложности
Модель глубокого обучения  YOLOv5	Реконструированные MPT изображения мозга, 400 образцов, 2 класса: доброкачественные и злокачественные	классификация	95,3% Точность  95,17% Достоверность  94,98% Полнота  95,23% Специфичность	Подходит для реального применения

**Инструменты, технологии и модели ИИ, используемые для обнаружения  
и дифференциальной диагностики опухолей мозга**

Модель	Датасет	Функция	Производительность	Примечание
ResNet-50 v2	Датасет ВТ	классификация	99,68% Точность	Ограниченное применение из-за высокой сложности
Автоматизированный детектор опухолей	Датасет ВТ	только сегментация	95% Точность	Необычная архитектура, но неустойчивая производительность при разных датасетах
Сверточная нейросеть и улучшенный алгоритм «поиска воробьёв»	Датасет ВТ	только сегментация	94,77% Точность 97,15% Полнота 67,16% Специфичность	Ограниченное применение из-за высокой сложности

ним относится и термин *дополненный интеллект / augmented intelligence*. Дополненный интеллект также известен как *усиление интеллекта / intelligence amplification IA*, а также *когнитивное усиление / cognitive augmentation*. Дополненный интеллект – это форма машинного обучения, которая направлена на улучшение принятия решений человеком и связанных с этим действий. Он использует такие технологии:

- Машинное обучение: системы обучаются на основе данных, выявляют закономерности и принимают независимые решения для обучения и совершенствования.
- Логическое рассуждение: использует логические методы для рационализации и получения выводов и заключений на основе имеющихся данных для решения проблем и составления прогнозов.
- Пространственная навигация: искусственные нейронные сети работают по принципу человеческого мозга.
- Обработка естественного языка: эта технология позволяет научить машины понимать и интерпретировать человеческий язык.
- Машинное зрение: позволяет компьютеру осуществлять автоматический анализ и проверку.
- Распознавание образов: это автоматическое распознавание образов в данных путем выявления закономерностей.

Дополненный интеллект реализует 5 основных функций:

1. Понимание: данные поступают в систему, а затем на их основе извлекается значение.
2. Интерпретация: по мере поступления новых данных старые данные анализируются и определяются новые наборы данных.
3. Причина-следствие: Новый набор данных затем определяет результаты.

4. Обучение: Человек предоставляет обратную связь по результатам или выходным данным, и система корректируется.

5. Гарантия: для обеспечения соответствия требованиям безопасности при управлении рисками используются технологии блокчейна.

Возможности для улучшения человеческого познания огромны: от поддержки неопытных врачей в выполнении сложных задач, решение которых доступно только высококласным специалистам, до фильтрации потока рутинных пациентов, чтобы врачи смогли уделять больше внимания нестандартным ситуациям. Кроме того, *IA* может помочь сократить количество врачебных ошибок, вызванных когнитивными искажениями, забывчивостью, фрустрацией или усталостью. При выполнении оперативных вмешательств *IA* может предложить некие особые возможности, которые не под силу рядовым хирургам. И всё-таки опрометчиво чересчур полагаться на автономно работающие платформы ИИ, которые могут допускать неточности в то время, как общественное мнение крайне нетерпимо к медицинским ошибкам и просчетам, а законодательное регулирование ИИ в здравоохранении ещё далеко от совершенства. Иначе это может привести к «третьей зиме ИИ» или в котловину разочарования / *trough of disillusionment* из ажиотажного цикла Гартнера / *Gartner Hype Cycle*. Осторожный подход к автономно функционирующим системам медицинского ИИ особенно актуален для потребительских нейросетевых приложений, поскольку они нормативно регулируются как развлекательные платформы или бытовые приложения, а их стандарты и контроль качества обычно весьма поверхностны. Негативные последствия здесь могут быть очень масштабными, учитывая широкое распространение потребительских приложений ИИ в здравоохранении и сложность в отсутствии экспертного надзора фиксации случаев нанесения ущерба от их использования.

Тут уместно вспомнить смешной, но мудрый каламбур о том, заменит ли ИИ когда-нибудь врачей-рентгенологов. Ответ на этот вопрос – нет; однако те рентгенологи, которые используют ИИ, скоро заменят рентгенологов, которые его не используют. Чтобы развивать и поддерживать ИИ в здравоохранении, нам необходимо широкомасштабное, всеобъемлющее углубление соответствующего направления профессионального образования, ориентированного на науку о больших данных / *Big data science* и эмоциональный ИИ / *Emotion AI*. Это расширение должно быть междисциплинарным и вовлекать разработчиков нейросетевых приложений, руководителей системы здравоохранения, передовых клиницистов, медицинские профессиональные ассоциации, неравнодушных пациентов, поскольку каждый из них может привнести свой жизненный опыт, а прогресс ИИ зависит именно от знающих людей, ответственно принимающих решения, которые уравнивают противоречивые мнения относительно простоты или



сложности внедрения различных инноваций, понимая сильные и слабые стороны ИИ. Чтобы начать решать проблемы *Emotion AI* Массачусетский технологический институт, университеты Гарвард, Стэнфорда и Техаса ввели у себя обучающие программы по цифровой этике. Ранее работавший в Google преподаватель компьютерных наук Стэнфордского университета М. Sahami высказался так: «Технологии не нейтральны. Выбор, который делается при создании технологий, затем имеет социальные последствия» (Singer, 2018) [149].

Вот некоторые важные темы в области ИИ, которые следует включить в обучающие программы непрерывного медицинского образования: 1) оценка необходимости, обоснованности и применимости алгоритмов медицинского ИИ; 2) понятие алгоритмической производительности и её влияние на последующее клиническое использование; 3) институциональная ответственность ИИ, её влияние на индивидуальную ответственность медицинских работников и врачебные ошибки; 4) вопросы стандартизации и прозрачности инструментов ИИ; 5) новые технологии в области ИИ, их использование и их зависимость от данных пациента и отношений врач-больной; 6) справедливость и инклюзивность ИИ; 7) вопросы консультационного взаимодействия с экспертами по ИИ. По мере развития области характер и акцент упомянутых тем, естественно, будут меняться, что потребует их периодического пересмотра и обновления. Итак, профессиональное медицинское образование должно включать в себя то, как критически оценивать полезность и риск инструментов ИИ в клинической практике. Учебные программы должны обеспечивать понимание того, как разрабатываются модели ИИ, критерии и соображения по использованию инструментов ИИ, как лучше всего задействовать такие инструменты, уделяя приоритетное внимание потребностям пациентов, и в каких случаях человеческий надзор необходим [11, 33].

Текущие программы профессиональной подготовки врачей несут на себе бремя растущих объемов научных знаний, которые обучающийся должен освоить в ограниченном временном отрезке. Профессорско-преподавательский состав признаёт непрактичность постановки вопроса о том, чтобы каждый врач стал экспертом в области ИИ. Гораздо лучше, если упор будет сделан на формировании мультидисциплинарных бригад, где каждый член команды кроме своих специальных знаний будет обладать базовыми знаниями по ИИ и обращаться за консультацией к разработчикам платформы по мере необходимости. Такие консультации можно проводить виртуально, поддерживая работу мультидисциплинарной бригады и групповое принятие решений. Это будет гораздо дешевле, чем если бы программисты присутствовали оффлайн. Региональные или контент-экспертные консультации можно было бы использовать во многих учреждениях здравоохранения. Примером консультационных центров государственного

масштаба является финансируемая Национальным институтом здравоохранения США *Сеть недиагностированных заболеваний / Undiagnosed Diseases Network UDN*, которая стремится улучшать и ускорить диагностику редких и трудно диагностированных патологий. По идее следует шире использовать дополнительные программы переподготовки для целевых категорий медицинских специалистов, которые будут наиболее психологически готовы к освоению требующихся при эксплуатации ИИ наборов навыков. Поэтому маловероятно, что слишком уж большое количество рабочих мест в сфере здравоохранения будет потеряно, хотя всё же следует ожидать несоответствия теоретических знаний и практических цифровых навыков [98].

Производителям медицинского ИИ в первую очередь нужно выводить на рынок интегрированные инновационные фреймворки для обслуживания нейросетевых платформ, в которых во взаимодействии человека с компьютером сбалансированы этическая инклюзивность и передовые стандарты в разработке программного обеспечения. Эти фреймворки должны учитывать контекст обучающейся системы здравоохранения и быть нацелены на решение конкретных задач. Важно подходить к медицинскому ИИ как к одному из многих инструментов поддержания здоровья и благополучия населения. Его следует использовать для реальных проблем, в решении которых более простые и дешёвые инструменты использовать непрактично. Платформы ИИ имеют достаточно высокую стоимость, выходящую за рамки возможностей медицинских организаций. Система оказания медицинской помощи в значительной степени сосредоточена на соматической патологии человека, и практически не учитывает социальные детерминанты здоровья и психосоциальные факторы риска. Алгоритмы ИИ имеют потенциал при достижении соответствующего консенсуса интегрировать личные и публичные данные людей для по-настоящему персонализированного здравоохранения.

ИИ мог бы пригодиться там, где планируются масштабные и коренные преобразования, а ресурсов не хватает даже для текущих потребностей. Такие проекты могут предусматривать реорганизацию помощи наиболее сложным пациентам с множественными сопутствующими хроническими заболеваниями, а также пожилым мало обеспеченным людям, проживающим в плохих условиях. Телемедицинский ИИ поможет при ликвидации последствий стихийных бедствий в сельских районах, когда ресурсы здравоохранения ограничены и доступ населения к ним затруднен из-за отдаленности, поэтому сортировка и автоматизированное распределение ресурсов могут быть реализованы с помощью алгоритмов ИИ. Мы настоятельно рекомендуем, чтобы надежная и зрелая система управления информационными технологиями была заранее внедрена в учреждении здравоохранения до начала работ по развертыванию платформы ИИ. Развитая аппаратная инфраструктура, управление изменениями, инклюзивное взаимодействие со всеми

стейкхолдерами и мониторинг безопасности требуют вложения существенного объема финансовых средств. Медицинским организациям, не обладающим такими ресурсами и компонентами инфраструктуры, надлежит обзавестись ими перед внедрением приложений ИИ.

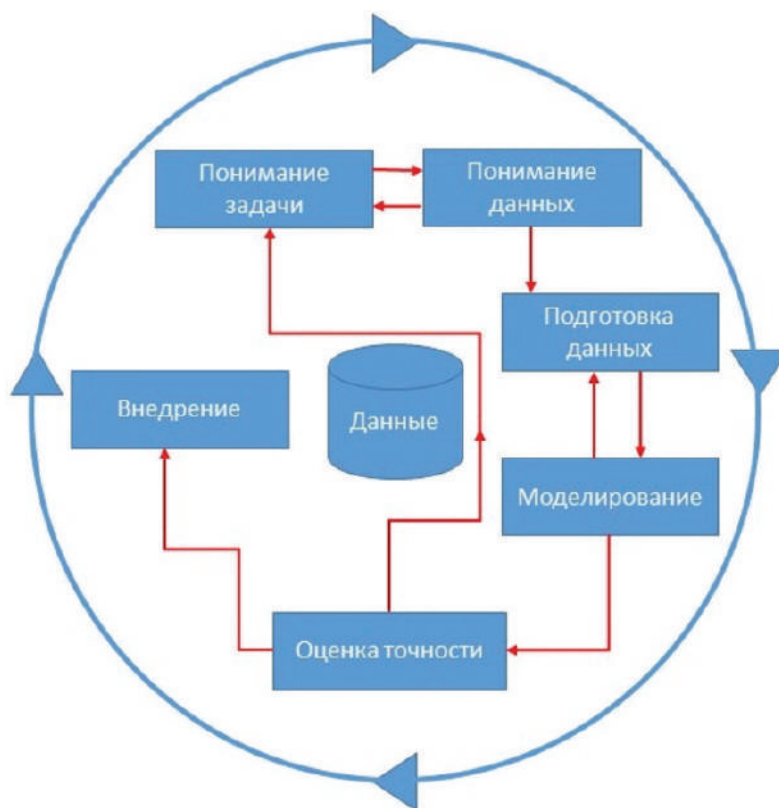
Авторы также рекомендуют пошаговый подход к регулированию ИИ с учетом уровня риска для пациента, уровня автономности ИИ и того, насколько функциональным окажется нейросетевой алгоритм. Поскольку модели *ML* постоянно обучаются на новых данных, регулирующие органы здравоохранения должны предпринимать меры эксплуатационного надзора для обеспечения бесперебойной высококачественной работы моделей. Связи между инновациями, безопасностью, прогрессом и регулированием очень сложны. Регуляторы должны участвовать в совместных усилиях с заинтересованными сторонами и экспертами по мониторингу эксплуатируемого медицинского ИИ на предмет оценки эффективности и безопасности. При этом прозрачность алгоритма помогает в предоставлении хорошо проверенных решений. Чтобы обеспечить как разработку ИИ, так и надзор за ним, правительственные учреждения должны инвестировать в инфраструктуру, которая способствовала бы более широкому доступу к источникам данных, справедливости и защите информации.

Реализация проектов с применением технологий ИИ достаточно типовая. Разработчики интеллектуальных систем, как правило, придерживаются методологии *CRISP-DM Cross Industry Standard Process for Data Mining* / межотраслевой стандартный процесс для исследования данных (рис. 11).

Основными этапами методологии являются: понимание бизнес-задачи; понимание данных; подготовка данных; моделирование; оценка точности модели; внедрение в бизнес-процесс. Рассмотрим основные факторы успеха проекта на примере предиктивного обслуживания оборудования.

#### 1. Постановка целей и формализация задачи

Для начала необходимо разобраться с приоритетами, сформулировать задачу и цели проекта. Без понимания, что делаем и с какой целью, невозможно выбрать показатели успеха проекта и составить техническое задание / ТЗ. Основной целью внедрения систем предиктивного обслуживания ИИ является сокращение издержек из-за простоя оборудования и внеплановых ремонтов. Целесообразно начинать с выбора наиболее критических для процесса оказания медицинской помощи объектов или узлов, начиная с компьютерного оборудования и кончая программным обеспечением. Требование по количеству отказов продиктовано не только эффективностью инвестиций, но и особенностями построения прогнозных моделей. Например, если в исторических данных количество отказов не больше 5, то требование точности предсказаний отказов на уровне 95% будет завышенным – модель ИИ не сможет обучиться до указанного уровня точности на таком количестве данных. Для



**Рис. 11.** Этапы построения решения ИИ в области анализа данных.

подобных случаев задача может быть переформулирована с прогнозирования отказов на задачу выявления отклонений в работе нейросети.

Для снижения риска рекомендуется начинать внедрение систем предиктивного обслуживания с реализации пилотных проектов. Главная задача – подтвердить или опровергнуть гипотезу, выявить ограничения и дополнительные требования для улучшения качества решения. Например, может оказаться, что на точность прогнозирования влияют факторы, данные по которым не собирались, и при текущей точности внедрение в промышленную эксплуатацию нецелесообразно.

В процессе пилотного тестирования происходят: передача массива данных модели; обучение модели; тестирование модели на исторических данных или в полевом эксперименте. На каждом этапе важно обозначить состав работ, сроки проведения и целевые показатели. Примерные сроки реализации составляют от 2-х до 3-х месяцев.

## 2. Наличие данных для построения модели ИИ

Чем больше данных и чем лучше их качество, тем выше уровень модели, поэтому важно накапливать данные, чтобы глубина хранения позволяла строить качественные предсказательные модели с высоким уровнем точности. Подготовка данных занимает 80% времени в проектах анализа данных и является самым трудоемким и ответственным этапом в разработке. Одна из глобальных стратегических



инициатив, рекомендуемая к внедрению, это практика управления данными *Data Governance*, призванная помочь медицинской организации повысить качество корпоративных данных, увеличить их ценность и монетизировать в дальнейшем. Набор данных зависит от решаемой задачи, но обязательное требование – это наличие исторических данных в контексте заданной темы. Для того чтобы внедрить систему предиктивного обслуживания в производственный процесс учреждения здравоохранения, необходимо провести аудит существующих данных, оцифровать те, которые хранятся в бумажных носителях, и наладить загрузку в базы данных на платформу в реальном времени.

### 3. Оценка результата и эффективности решения

Точность построенной модели на текущих данных может быть неудовлетворительной и пользы не принесет. Важно перед принятием решения о внедрении просчитать экономическую выгоду дальнейшей реализации. Возможно, потребуется накопление дополнительных данных для улучшения качества модели. Результатом пилотного проекта и моделирования прогнозирования ошибок и сбоев будут варианты управленческих решений:

- о внедрении платформы ИИ в эксплуатацию, если пилот успешен согласно установленным в ТЗ критериям;
- о продлении пилота на определенный срок, если времени оказалось недостаточно, но промежуточные итоги удовлетворительные;
- о приостановке, если требуется обогатить датасет дополнительными данными;
- об отказе от внедрения, если пилот завершился неудачно согласно установленным в ТЗ критериям с анализом причин неудачи и выводами о дальнейшей судьбе инициативы.

Сама прогнозная модель законченным решением не является, это необходимое ядро системы. Для того чтобы внедрить предиктивное обслуживание в лечебно-диагностический процесс, необходимо, как уже говорилось, наладить загрузку данных в систему предиктивного обслуживания в реальном времени. Тогда модель заработает и начнет формировать ответы в зависимости от поставленной задачи. Частным примером реализации системы может быть создание веб-портала с разграниченными правами пользователей. Дашборды пользователей позволяют просматривать: список по контролируемой инфраструктуре и информацию о её состоянии; данные по состоянию элементов в эксплуатационной карточке; список обнаруженных ошибок и сбоев; прогноз по их последствиям. Кроме того, должна быть предусмотрена рассылка уведомлений ответственным пользователям о событиях, связанных с нарушениями в работе алгоритма.

Если принято решение о внедрении в эксплуатацию, рекомендуется составлять: описание требований к технологическому обеспечению, полный перечень используемых данных, шаблоны выходных документов и отчетов, которые должна формировать модель ИИ, архитектуру платформы, требования к показателям её

надежности, установленные ограничения и проч. Список может быть расширен или сокращен в зависимости от видения ведущих стейкхолдеров, но подобные документы позволяют снижать риски внедрения и подсвечивают слепые пятна, с которыми лучше разобратся до начала развертывания проекта. Полностью автономная эксплуатация платформ ИИ еще впереди. Однако уже сегодня они могут служить вспомогательным инструментом медицинскому персоналу, помогая врачам принимать обоснованные решения. Благодаря раннему предупреждению о потенциальных отказах и проблемах уменьшается количество дефектов оказания медицинской помощи. Возможность удаленного мониторинга и ускорение процесса анализа ошибок повышают эффективность производственного процесса учреждения здравоохранения, а профилактика сбоев оптимизирует рабочие регламенты и позволяет увеличивать прибыль.

#### **4. ЭМОЦИОНАЛЬНЫЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ**

##### **Эмоциональный интеллект в деятельности врача**

Эмоциональный интеллект (ЭИ) – это умение понимать и управлять своими эмоциями, а также умение понимать и управлять чужими эмоциями, вследствие чего – выстраивать долгосрочные отношения с другими людьми. Профессиональное самоосознание врача в рамках системы эмоционального интеллекта – это умение распознавать свои эмоции, эмоциональные импульсы и причины их возникновения, а также уверенность в себе: независимость в принятии решения и адекватная оценка собственных способностей. Каким образом профессиональное самосознание как компонент эмоционального интеллекта помогает в работе врача? К врачу обращаются, когда есть проблема, а значит, каждый рабочий день он, общаясь с пациентами, получает от них заряд негативных эмоций: страх, паника, печаль, беспокойство, отчаяние. В такой ситуации врачу и самому может понадобиться психологическая помощь. Чтобы этого не произошло, врачу необходимо профессиональное самосознание, то есть умение определять, какие именно эмоции он испытывает, причины их возникновения – почему именно он находится сейчас в этой эмоции, может быть, он как бы заразился ею от своего пациента. Каждому врачу нужно научиться понимать себя, отслеживать свои эмоции и чувства, ведь, как известно, ни одна специальность не приносит порой столько моральных переживаний, как врачебная.

Формирование модели взаимоотношения врача и пациента не является универсальным процессом и зависит от многих факторов (Абаева О.П. с соавт. 2019; Романова Т.Е. с соавт. Терапевтический архив, 2024) [1, 25]. Нередки случаи, когда врач заражается эмоциями пациента, начинает излишне ему сопереживать, даже

можно сказать, разделяет с ним его чувства и попадает под его влияние. И начинает терять свою независимость в принятии решения о методе и процессе лечения. Таким образом, складывается ситуация, что пациент начинает манипулировать врачом, объясняя ему как правильно надо лечить данное заболевание. Иногда даже в жесткой и грубой манере давит на врача, заявляя, что он, больной, уж точно знает, в чем именно кроется проблема. Поэтому так важна саморегуляция врача, то есть умение управлять своими эмоциональными импульсами, справляться со стрессом и переводить себя в ресурсное состояние, в котором принимать максимально взвешенные и объективные решения. Как уже было сказано выше, в течение дня врач получает огромное количество негативных эмоций от пациентов. Не будем забывать, что врач не робот, а живой человек со своими жизненными трудностями, семейными проблемами, индивидуальными особенностями и характером. Врач, так же, как и все, испытывает всю палитру эмоций: от позитивных до негативных. Хорошее настроение надо уметь себе создавать. Принцип эмоционального интеллекта в том, что мы можем быть счастливыми или несчастными по своей воле. Никто не сделает нас счастливым, кроме нас самих. Эмоции влияют на то, какие именно решения мы принимаем в работе и личной жизни, на качество этих решений. Врач как никто другой должен принимать решения, основанные на фактах и здравом смысле, а не на эмоциональных порывах (Павлов С.В. с соавт., 2019) [18].

Мотивация врача – сохранение и поддержание интереса и вовлеченности в профессию, а также получение удовольствия от содержания и результата работы. Эмоциональное выгорание, а также потеря мотивации врачом – наиболее частные проблемы любого профессионала, задействованного в работе с людьми (Reshetnikov A.V. et al, 2024). Любой врач подвержен двум крайним эмоциональным состояниям в работе с пациентами: проживать боль, страдания и жизненный путь своего пациента. Переживать эмоции и чувства пациента как свои собственные, брать на себя полную ответственность за жизнь другого человека. И брать всю вину на себя, если процесс лечения закончился трагически. Деперсонализация – формальное и безразличное выполнение своего профессионального долга. В крайних случаях – циничное отношение к людям и эмоциональная отстраненность. В первом случае такое отношение быстро приведет к физическому и эмоциональному истощению, что неизбежно скажется на личной жизни врача и его собственном психическом здоровье. Во втором случае поверхностное отношение к работе, пессимистичный настрой в процессе лечения, отсутствие потребности профессионального развития приведут к частным врачебным ошибкам. Оба варианта медленно и верно убьют любовь к профессии. Врачу необходимо научиться быть эмоционально сбалансированным, чтобы сохранять адекватное восприятие реальности, что особенно актуально в условиях повышения на-

грузки на медицинских работников (Хазов М.В. с соавт. 2015). Безусловно, вопрос сохранения внутренней мотивации врача и любви к работе всегда индивидуален, и каждый случай эмоционального выгорания должен рассматриваться исходя из конкретной ситуации (Бердутин В.А., Романова Т.Е., 2023) [9, 36, 137].

Фундаментальная часть практики врача основана на лидерских качествах, особенно с учетом острой потребности в квалифицированных лидерах для продвижения и управления результатами, ориентированными на пациента и медицинскую организацию. Знание того, как использовать эмоциональный интеллект, может улучшить отношения врача с пациентами и взаимодействие с коллегами. В связи с растущим вниманием к медицинской специализации и новыми инициативами в области медицинского обслуживания потребуется больше врачей-лидеров. Термин эмоциональный интеллект относится к совокупности способностей, талантов и компетенций, которые влияют на способность человека добиться успеха на фоне требований, проблем и стресс-факторов окружающей среды. Этот термин стал революционным в менеджменте и трансформировал лидерство, введя идею о том, что осознание своих эмоций необходимо для преодоления стресса. ЭИ, называемый также *эмоциональным коэффициентом EQ*, является фундаментальным этапом на пути к успеху медицинских работников, в частности, для преодоления сложностей, присущих современной системе здравоохранения.

Эмоциональная грамотность имеет решающее значение в управлении процветающим коллективом. В отличие от IQ, который используют для оценки интеллекта с помощью стандартных тестов, ЭИ – это то, что каждый член социума обязан развить в себе с рождения, чтобы не выглядеть аутистом. В действительности повышение эмоционального интеллекта должно быть постоянной профессиональной и личной задачей, особенно для руководителей учреждений здравоохранения. Следует отметить, что 80–90 % навыков, которые отличают высокоэффективных сотрудников от неэффективных, относятся к категории ЭИ. В цифровую эпоху развитый ЭИ стал насущной потребностью межличностного и делового взаимодействия в практиках не только реального бытия социума, но и его виртуального бытия, включая виртуальное лидерство – одно из последних нововведений эмоционального ИИ *Emotion AI/EAI* (Мамина Р.И., 2023) [17].

Директор Института перспективных исследований мозга МГУ им. М.В. Ломоносова академик К.В. Анохин под термином сознание подразумевает информационный трафик, который проходит по структуре нейронной гиперсети мозга. Образно говоря, разум можно представить как дорожную сеть, сознание как трафик, а эмоции как интенсивность трафика. М. Raichle выделяет следующие сети мозга: сеть пассивной работы или *дефолт система / default mode network* мозга, сеть выявления значимости и центральная исполнительская сеть, поэтому предоставленный самому себе человеческий мозг естественным образом включается в размышления



о социальных отношениях. Сеть выявления значимости проводит субъективную оценку образов, мыслей и переживаний; центральная исполнительская сеть отвечает за потребление информации. Представляющая особый интерес дефолт система мозга обеспечивает взаимодействие отделов головного мозга в моменты, когда человек находится в состоянии пассивного бодрствования и не занят какой-либо познавательной деятельностью. Специалисты хорошо знают, что наивысшую активность мозг проявляет в режиме *блуждающего ума*. Нейронная сеть оперативного покоя выполняет следующие функции: обработка информации о самом себе, представление и размышление о других людях, моральные рассуждения, воспоминания прошлого и размышления о будущем, понимание истории. Таким образом, дефолт система, позволяя мозгу работать в автономном режиме или в режиме покоя, осуществляет анализ и классификацию прошедших событий, а также формирует субъективные отношения к данным событиям, то есть, другими словами, подсознательно формирует эмоции личности.

Итак, интерпретация эмоций, а также умение воздействовать на эмоции других людей составляет эмоциональный интеллект, который является неотъемлемой частью общего интеллекта личности. Естественным образом возникает вопрос: нужен ли искусственному интеллекту эмоциональный интеллект? Однозначно утвердительного ответа на данный вопрос пока не дал никто. С одной стороны, очень хочется сделать компьютерный разум более человечным, по крайней мере, способным имитировать и распознавать эмоции; с другой стороны, весьма опасно позволять искусственному интеллекту по своему усмотрению воздействовать на эмоции людей. Поэтому никогда нельзя забывать о том, что на протяжении всей многовековой истории человечества люди всегда были гораздо более успешны в изобретении инструментов и устройств, чем в их разумном использовании на благо, а не во вред самим себе и обществу в целом.

### **Эмоциональные вычисления**

Созданием технологий, ответственных за обработку эмоциональной информации в системах ИИ, занимается направление, получившее название *эмоциональные вычисления* / *affective computing*, что равнозначно термину эмоциональный искусственный интеллект. Считается, что направление появилось в 1995 году с выходом в свет работы R.W. Picard из Массачусетского технологического университета. В своей книге «Affective computing» она впервые показала важность эмоций не только в реальной, но и в виртуальной среде, а также определила возможности распознавания и моделирования эмоций с помощью компьютерных систем. Обозначившееся новое направление исследований эмоций, посвященное тематике ЭИ в контексте с ИИ, послужило стимулом для дальнейшего развития направления взаимодействия человека с компьютером. Профессор R.W. Picard

предложила модели для компьютерного распознавания человеческих эмоций, обосновав их практическое применение. На заре этого направления доминировали различные классические техники обработки сигнала, такие как спектральные и кепстральные разложения вкупе с регрессионными деревьями, марковскими моделями и т.д. Кепстр / *cepstrum* – один из видов гомоморфной обработки сигналов, функция обратного преобразования Фурье от логарифма спектра мощности сигнала. В эпоху революции глубокого обучения в арсенал разработчиков систем для анализа эмоций добавились многослойные нейронные сети. Модели, предназначенные для распознавания эмоций, в наши дни обычно представляют собой рекуррентные, свёрточные или свёрточно-рекуррентные нейронные сети. А после 2017-го года к ним добавились ещё и трансформеры. Причём стоит заметить, что характерными маркерами эмоций могут быть не только лингвистические показатели, т.е. смыслы слов и выражений, но и экстралингвистические, такие как тон голоса, интонации, выражение лица, динамика тела и жестикуляция.

Эмоциональная информация рассредоточена по разным каналам: мы можем найти её и в звуке, и в видео. А если рассматривать сенсорную сферу, проявление эмоций можно найти в касаниях, и это тоже может быть важно для роботов. Ещё один канал, в котором эмоции проявляются на уровне активности мозга – энцефалографический. Его можно использовать, например, при проектировании детектора лжи со встроенным ИИ. Одним из подразделов эмоциональные вычисления является *сентимент-анализ* / *sentiment analysis*, цель которого состоит в определении смысловой окраски высказывания: негативная, нейтральная или позитивная. В наши дни эту задачу решают при помощи трансформерных моделей, подобных *BERT Bidirectional Encoder Representations from Transformers* / *двухнаправленные презентации кодировщика для трансформеров*. BERT представляет собой нейронную сеть, основу которой составляет композиция кодировщиков трансформера. Она сама является автокодировщиком. В данной модели используется механизм *self-attention*, позволяющий взаимодействовать с другими словами в предложении, что сделало возможным по-новому взглянуть на проблему обучения ИИ.

Давайте рассмотрим поподробнее один из каналов, которые задействуются при общении человека с машиной, например, аудиальный канал. Распознавание эмоций в речевом канале – это одна из наиболее распространенных задач в области эмоционального ИИ. Чаще всего для построения модели применяются глубокие сети, которым на вход подаются различные представления звукового сигнала: спектрограммы, хромограммы, последовательности наборов мел-кепстральных коэффициентов и т.п. Такие модели решают задачу классификации или регрессии. Чтобы обучить модель, распознающую эмоциональную окраску речи, нужно подготовить обучающую выборку. А для этого нужно условиться, какое представление эмоций мы будем использовать. Возможные

классификации предоставляет язык разметки EmotionML 1.0. Он содержит несколько «эмоциональных» словарей, основанных на научных классификациях, среди которых большая шестёрка эмоций американца Пола Экмана / *P. Ekman*: отвращение, печаль, гнев, страх, счастье и удивление. Другой эмоциональный словарь, предусмотренный EmotionML 1.0, основан на концепции соответствия эмоций тенденциям действия *action tendencies*, разработанной голландцем Нико Фрейдой / *N.H. Frijda*. Этот словарь включает в себя 12 эмоций: безразличие, высокомерие, гнев, желание, интерес, наслаждение, отвращение, покорность, смирение, страх, удивление и шок.

Не стоит простодушно считать, что авторы эмоциональных словарей как-то соревновались друг с другом в составлении списков эмоций. В основе больших эмоциональных словарей обычно лежит анализ лингвистических данных: статистики употребления слов, используемых для передачи эмоционального фона речи. При этом сами словари являются своего рода побочным продуктом исследований по построению некоего эмоционального пространства, в котором каждая эмоция представлена несколькими независимыми друг от друга компонентами. Одну из попыток построить такое пространство предпринял Джеймс Рассел / *J.A. Russell*, разложив эмоции по двум шкалам: 1) шкала *удовольствие-неудовольствие* характеризует позитивный или негативный характер эмоции. 2) Шкала *возбуждение-сон* характеризует активность или пассивность психического состояния. Эта работа вызвала закономерную критику, поскольку мир эмоций не сводим к двумерному пространству. Поэтому двухмерный формат модели было предложено дополнить, представив её в виде сетки или решётки GRID.

Уже при наличии эмоционального континуума мы вынуждены вместо задачи классификации, когда мы имели дело лишь с несколькими классами эмоций, переходить к решению задачи регрессии. Теперь от модели требуется не столько предсказание метки конкретного эмоционального класса в соответствии с выбранным эмоциональным словарём, сколько оценка величины каждой из выбранных компонент эмоции. Для этой цели в стандарт EmotionML 1.0 встроена процедура измерения эмоций. Кроме упомянутого выше формата GRID с 4-мя шкалами, стандартом предусмотрена возможность использования пространства PAD удовольствие-возбуждение-доминирование / *pleasure, arousal, dominance*, основанного на 3-ёх соответствующих шкалах, а также плоской шкалы интенсивности эмоции. Кстати, модели эмоций могут быть и мультимодальными, так как при оценке эмоциональной окраски речи люди неизбежно ориентируются не только на звуковые признаки, но и на содержание сказанного. А если используется видеозапись выступления человека, то к числу признаков добавится ещё и видео синхронизация с выражением лица и движениями тела говорящего. В данном случае нужна модель, которая может обрабатывать мультимодальную информацию.

Разметка данных всегда являлась проблемой из-за высокой субъективности оценок эмоциональной окраски речи. Допустим, один человек слышит в печаль в какой-то фразе, а другому фраза кажется буднично нейтральной. Если пользоваться системой шкал, то в датасете полученные от разных людей значения оценок, по каждой шкале можно усреднять. При использовании же словарей придётся либо доверять большинству оценщиков, либо выбраковывать получившие неоднозначные оценки фразы. На момент 2009-го года стало известно о существовании порядка сотни эмоциональных датасетов. Однако таких же объёмных наборов данных для оценки эмоциональной речи, какими обладают фирмы ImageNet и LibriSpeech, в публичном доступе до сих пор нет. Вот некоторые наиболее популярные на сегодняшний день публичные датасеты эмоциональной речи:

RAVDESS состоит из записей 24 профессиональных актёров: 12 мужчин и 12 женщин, озвучивающих 2 фразы «у двери разговаривают дети», «собаки сидят у двери» на английском языке с североамериканским акцентом в двух вариантах: речь и пение, по две озвучки на каждый вариант. В качестве эмоционального словаря разметки использована *большая шестёрка* эмоций, к которой было добавлено *спокойствие*. Каждая фраза представлена в датасете двумя уровнями эмоциональной интенсивности для каждой из эмоций, а также однократно с нейтральной окраской. Каждая запись присутствует в датасете в 3-ёх модальностях: только видео, только звук, звук вместе с видео. RAVDESS считается одним из наиболее качественных датасетов эмоциональной речи, но лексически он крайне беден.

SAVEE состоит из записей 4-ёх актёров-мужчин, говорящих на родном для них британском английском. В качестве эмоционального словаря снова выбрана большая шестёрка, при этом фразы с нейтральной эмоциональной окраской записывались дважды. Сами фразы были выбраны из корпуса TIMIT (датасет с записями 630 дикторов), для каждой эмоции было взято 15 фраз, при этом из них 3 были общими для всех эмоций, 10 – разными для разных эмоций, но без эмоциональной специфики, а ещё 2 фразы были основаны на текстах, имеющих специфическую эмоциональную окраску для данной эмоции, например, «Кто одобрил счёт с неограниченным расходным лимитом?» для гневной эмоции. К сожалению, объём этого датасета крайне мал, что создаёт проблемы для разработчиков.

SEMAINE – это аудиовизуальная база данных, ставшая одним из продуктов исследовательской программы по созданию аудиовизуальной диалоговой системы, способной вовлечь человека в длительный эмоционально окрашенный разговор, *SAL sensitive artificial listener / чувствующего искусственного слушателя*. По сути, разговор с агентом SAL для человека напоминает обычный разговор при помощи системы видеосвязи с той лишь разницей, что собеседником является виртуальный персонаж, внешний облик которого (лицо, мимика, движения губ во время речи) в ре-



альном времени генерируется при помощи библиотеки для трёхмерной визуализации. Данные, содержащиеся в базе SEMAINE, были получены в результате взаимодействия между пользователями и человеком-оператором, имитирующим чувствующего и кушённого слушателя, а затем и ассистентом на базе нейросетевой модели. База включает записи 959 диалогов, в которых участвовало 150 человек. Длина каждой записи составляет около 5 минут. Все диалоги были расшифрованы и размечены при помощи эмоциональных меток, где использовалась система с пятью шкалами и 27 эмоциональными классами. Для части записей присутствует разметка при помощи *FACS / системы кодирования лицевых движений*. Используя FACS, можно с лёгкостью отличить, например, дежурную американскую улыбку от искренней улыбки. Недостатки этого датасета: различные эмоции представлены крайне неравномерно, не сбалансированы ни состав участников исследования, ни лексическая основа диалогов. Достоинством же датасета является уникальная детальность разметки.

TESS – весьма скромный по размерам датасет, в котором используемый набор фраз состоит из так называемой фразы-носителя «Скажи слово...» и списка из 200 добавляемых к ней слов. Каждая из фраз произносилась 2-мя актрисами 26 и 64 лет с 7-ью различными типами эмоциональной окраски. Таким образом, в сумме было получено  $200 \times 7 \times 2 = 2\,800$  записей.

ЕМО-DB – немецкоязычный массив данных, который долгое время пользовался большой популярностью у исследователей эмоциональной речи. Десять актёров (5 женщин и 5 мужчин) имитировали эмоции, произнося по 10 предложений (5 коротких и 5 более длинных), относящихся к повседневному лексикону. Помимо звука были записаны электроглоттограммы. Электроглоттография основана на измерении динамики электрического сопротивления гортани во время произнесения фраз, что достигается при помощи пары электродов, располагаемых на передней поверхности шеи по обе стороны щитовидного хряща. 10 актёров  $\times$  10 предложений  $\times$  7 эмоций, включая нейтральную, составили 700 записей. Плюс часть записей была выполнена повторно, поэтому в базе содержится на 100 записей больше. Все записи были подвергнуты оценке с привлечением 20 оценщиков. После этого в записях со средним уровнем узнавания эмоции более 80% и средней оценкой убедительности более 60% была дополнительно оценена интенсивность проявления эмоции.

ИЕМОСАР – это массив, созданный в лаборатории анализа и интерпретации речи университета Южной Калифорнии, включающий в себя записи спонтанных диалогов 10 участников, а также их разговоров на основе заранее подготовленных сценариев. Данные состоят из аудиозаписи с расшифровкой, видео, а также подробной информации о выражении лица и движениях рук, а также эмоциональной разметки, куда входят большая шестёрка + другая эмоция + нейтральная окраска, а также оценка эмоций по 3-ём шкалам: валентность, активация и доминирование.

RUSLANA – первая открытая русскоязычная база данных эмоциональной речи, содержащая записи 61 человека (12 мужчин и 49 женщин), которые произносили 10 предложений с выражением следующих эмоциональных состояний: удивление, счастье, гнев, грусть, страх и нейтрально без эмоциональной окраски. Таким образом, в сумме база содержит  $61 \times 10 \times 6 = 3\,660$  записей. Хотя с момента появления RUSLANA свет увидели ещё несколько открытых русскоязычных эмоциональных датасетов, например, аудиовизуальный RAMAS и весьма внушительный по объёму содержащий более 20 000 записей набор эмоциональной детской речи EmoChildRu, открытых датасетов взрослой эмоциональной речи, превосходящих RUSLANA по объёму, до сегодняшнего дня так и не создано.

Эмоциональные вычисления могут использоваться в полностью автоматизированных диалоговых медицинских системах, чтобы робот-оператор определённым образом реагировал на те или иные эмоциональные нотки в речи пациента. А чтобы получить эмпатичного бота, стремящегося в ходе общения научиться выбирать такие слова, которые максимизируют положительную реакцию собеседника, можно в скрипте робота предусмотреть несколько вариантов ответа, чтобы накапливать статистику эмоциональной реакции людей на каждый из этих вариантов, и потом использовать реплики, приводящие к более благоприятным реакциям. Эмоциональным может быть не только смысл и синтаксис фразы, эмоции могут быть у интонационной окраски произносимого. Были предложены модели, способные в режиме обучения без учителя выучивать для каждой фразы некоторые стилистические векторы. К числу таких моделей относятся такие модели, как *Tacotron TP-GST / Глобальные стилевые токены*, предсказанные на основе текста и *GMVAE-Tacotron / Вариационный автокодировщик* на основе смеси гауссовских распределений. Используя векторы, выученные моделью для фраз обучающей выборки, в качестве библиотеки стилей, можно добиться неплохой управляемости стилистикой синтеза. При этом отдельная модель может быть использована для того, чтобы построить стилистический вектор фразы на основе семантической информации. То есть, проще говоря, обучить модель, которая будет, исходя из смысла фразы, выбирать для неё правильную интонацию.

Моделирование эмоциональной окраски речи, а также имитация мимики в случае наличия визуального аватара, позволяет сделать бота уже не просто системой, обладающей «пассивным» эмоциональным интеллектом. Покинув границы этой парадигмы, можно замахнуться на то, чтобы активно влиять на эмоциональную сферу человека, формируя полноценное двустороннее эмоциональное взаимодействие. В здравоохранении самые амбициозные надежды связаны с тем, что исходя из эмоционального статуса пациентов ЭИИ поможет точно диагностировать некоторые психические заболевания, предотвращая эксцессы

деструктивные поведения пациентов. Особый интерес ЭИИ представляет для работы с больными, у которых имеются признаки деменции. Поскольку в период изоляции и низкой умственной активности болезнь быстро прогрессирует, то модели ЭИИ могут стать прекрасным вспомогательным средством для медперсонала по поддержке и уходу за такими пациентами. Израильская компания Beyond Verbal в сотрудничестве с клиникой Mayo Clinic путем фиксации в голосах пациентов вокальных биомаркеров научила модель ЭИИ прогнозировать не только болезни Альцгеймера и Паркинсона, но и скрытую сердечно-сосудистую патологию.

Эмоция – это не только субъективная оценка окружающего мира, но и стимул. От того, в каком состоянии находится человек, зависит его реакция на тот или иной стимул, его поведение. Имитируя эмоции, машина способна стимулировать у людей эмоции, например, имитируя гнев можно вызвать страх, а имитируя счастье – зависть. З. Фрейд предупреждал, что подавленные эмоции бесследно никуда не исчезают. Несмотря на то, что их заблокировали, подспудно они продолжают влиять на человека как бы изнутри. Отсюда можно заключить, что имитация эмоций искусственным интеллектом даёт ему возможность оказывать на пациентов как позитивное, так и негативное влияние. И, наделяя ЭИИ инструментами выявления значимости, отвечающими за оценку им своих решений, мы, по сути, способствуем формированию у него на основе субъективного опыта некоего внутреннего мира, то есть цифрового сознания.

Несмотря на бесспорные коммуникативные успехи роботизированных устройств со встроенным ЭИИ, у них присутствуют свои проблемные зоны. Например, эффектом *зловещей долины* называется явление, когда сильные эмоциональные реакции людей бывают вызваны антропоморфностью устройства. Человека повергает в шок дисбаланс между интеллектом умной машины и ее неказистым внешним видом или, наоборот, глупое поведение, демонстрируемое роботом-манекеном. Открытие эффект *зловещей долины* принадлежит известному японскому специалисту по робототехнике Масахиро Мори. Изучая, как человек реагирует на роботов, он обнаружил определенную закономерность: чем реалистичнее выглядит робот, тем выше эмоциональная оценка и чувство эмпатии; однако любые изъяны и отклонения в поведении похожего на человека робота моментально вызывают у людей реакцию отторжения. Если эти изъяны робота удаётся убрать, и он станет почти неотличимым от человека, то люди вновь начнут испытывать в отношении его положительные эмоции, что в свою очередь сказывается на поведении робота в позитивном ключе. Резкое снижение эмпатии, вызываемое эффектом *зловещей долины* считается одним из серьезнейших препятствий на пути взаимодействия человека и машины. Также наблюдается так называемый феномен *выученной предвзятости / AI bias*, который проявляется в дискриминации обученной на некачественных дата-

сетях платформы ЭИИ малообеспеченных и обделенных социальными благами людей.

Но вот что самое потрясающее – у эмоционального ИИ отсутствует стереотипное мышление. Робот хоть и распознает, но сам не испытывает эмоций, соответственно, ему не ведом внутренний конфликт. Поэтому, если его не научить плохому, то он не ответит грубостью на грубость, агрессией на агрессию, никогда не поддастся на провокации и, как следствие, сам будет служить для человека прекрасным примером терпимости, невозмутимости и беспристрастности. Отсюда понимание того, что значимость эмоционального ИИ для человека заключается не только в его коммуникативной полезности, но и, в том, что, взаимодействуя с такой технологией, эмоционально лабильные пациенты подсознательно будут ощущать себя совершенно спокойными и уравновешенными.

В настоящее время ЭИИ находится в процессе становления и развития. Главные акторы этого процесса – антропоморфные роботы, текстовые и голосовые чат-боты, а также видеоботы, которые активно демонстрируют полученные в ходе их обучения знания и навыки в области эмоциональной коммуникации. При этом в соответствии с ресурсами текущего ИИ эти знания и навыки развиваются и в процессе дополнительного обучения (с учителем, с частичным привлечением учителя, с подкреплением), и в процессе самообучения, которое стало возможным благодаря технологии генеративного ИИ, в частности, функционала новой нейросети *Chat-GPT*. И хотя эмоциональные компетенции современного ИИ все еще далеки от искомого совершенства, ученые предполагают, что рано или поздно настанет день, когда появятся системы, обладающие почти таким же уровнем креативности, ощущений и эмоционального интеллекта, что и человек.

### **Морально-правовой базис ЭИИ**

Поскольку технический прогресс неостановим, то появление обладающего киберсознанием, так называемого, сильного искусственного интеллекта / *artificial superintelligence* – всего лишь вопрос времени. Между тем создание такого ИИ, представляющего собой киберфизическую систему с искусственным самосознанием, способную формировать представление о себе и обучаться без учителя, создает немалые этические и правовые сложности. На повестке дня стоит выработка юридических норм, регулирующих отношения, связанные с использованием ЭИИ.

Несколько десятилетий назад знаменитый американский писатель-фантаст Айзек Азимов предложил 3 закона робототехники:

1. робот не должен наносить какой-либо вред человеческому здоровью, создавать угрозу жизни или своим бездействием допускать аналогичные последствия;

2. роботу вменяется в обязательном порядке исполнять приказы, отдаваемые человеком. Единственное исключение – отдаваемый



приказ не должен противоречить предыдущему положению;

3. роботу вменяется беспокоиться о собственной безопасности в той степени и в том объеме, в которых его действия не опровергают два предыдущих пункта.

Кстати, было бы неплохо конкретизировать первый закон, указав на запрет причинение вреда как физическому, так и ментальному здоровью человека. В этой связи в 2017 году исследовательским центром проблем регулирования робототехники и искусственного интеллекта *Робоправо* разработан проект под названием «Модельная конвенция о робототехнике и искусственном интеллекте», которая предусматривает обязанность соблюдения прав человека и общепринятых норм морали и нравственности. Сама по себе правовая база, регулирующая отношения в области ИИ состоит из международных актов и национальных нормативных актов. Окинавская хартия глобального информационного общества *Okinawa Charter on Global Information Society* от 22.07.2000 призвала правительства стран к принятию внутренних законов, касающихся ИИ. В 2017 г. был создан технический комитет по искусственному интеллекту *ISO/IEC JTC 1/SC 42 Artificial intelligence*. Рекомендации по искусственному интеллекту Организации экономического сотрудничества и развития ОЭСР *OECD / Council Recommendation on Artificial Intelligence* от 22.05.2019 содержат ряд стандартных требований, в том числе необходимость постоянной оценки и управления потенциальными рисками ИИ [26].

В Российской Федерации также приняты государственные нормативные акты, регулирующие использование искусственного интеллекта. В соответствии с указом Президента РФ от 10.10.2019 г. № 490 одним из основных принципов развития ИИ является безопасность: недопустимость использования искусственного интеллекта в целях умышленного причинения вреда гражданам и юридическим лицам, а также предупреждение и минимизация рисков возникновения негативных последствий использования технологий искусственного интеллекта. В соответствии с Федеральным законом от 24.04.2020 г. № 123-ФЗ в целях создания условий для разработки и внедрения технологий искусственного интеллекта установлен экспериментальный правовой режим в г. Москве. Важно отметить, что принципом специального режима является защита прав и свобод человека и гражданина, обеспечение безопасности личности, общества и государства [34, 35].

Кроме того, в соответствии с Федеральным законом «О техническом регулировании» от 27.12.2002 г. № 184-ФЗ необходимым условием любой продукции является ее безопасность — состояние, при котором отсутствует недопустимый риск, связанный с причинением вреда жизни или здоровью граждан, имуществу физических или юридических лиц, государственному или муниципальному имуществу, окружающей среде, жизни или здоровью животных и растений. Распоряжением Правительства РФ от 19.08.2020 г. № 2129-р утверждена Кон-

цепция развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники до 2024, в которой закреплена необходимость регуляторного воздействия, основанного на риск-ориентированном, междисциплинарном подходе и предусматривающем принятие ограничительных норм в случае, если применение технологий искусственного интеллекта и робототехники несет объективно высокий риск причинения вреда участникам общественных отношений, правам человека и интересам общества и государства [19, 20, 21].

Важно отметить, что, во-первых, в настоящее время отсутствует международный акт, закрепляющий универсальные принципы в области искусственного интеллекта, во-вторых, в законодательстве Российской Федерации в области искусственного интеллекта отсутствует понятие психического и морального вреда. Между тем 17.09.2020 г. были приняты Руководящие принципы по этике искусственного интеллекта *Ethics Guidelines for Trustworthy Artificial Intelligence*, в которых закреплено право человека на физическую, психическую и/или моральную неприкосновенность. В связи с этим в отечественном законодательстве или технических стандартах было предложено предусмотреть запреты на имитацию искусственным интеллектом некоторых видов эмоций, который могут причинить человеку физический, психический и моральный вред, например: гнев, печаль, беспокойство, страх, хамство и т.п., а также ограничить по времени имитацию искусственным интеллектом негативных эмоций при взаимодействии с человеком. Таким образом, наделение ИИ способностью выражать или имитировать эмоции при взаимодействии с человеком, должно осуществляться под неусыпным контролем экспертов в области коммуникативной психологии. Со всей очевидностью напрашивается вывод о том, что, с одной стороны, развитие ЭИИ, который является неотъемлемой частью технологии искусственного интеллекта, безусловно очень востребовано в отрасли здравоохранения, а, с другой стороны, это возможно лишь при условии наличия эффективно работающей правовой базы, направленной на защиту прав и свобод личности в области ИИ. Необходимо установить жесткие законодательные рамки и карательные меры, которые бы обеспечивали надежное ограничение использования ЭИИ вне правового поля [46, 69].

В целях обеспечения ускоренного развития искусственного интеллекта, проведения научных исследований в области искусственного интеллекта, повышения доступности информации и вычислительных ресурсов для пользователей, совершенствования системы подготовки кадров в России утверждена Национальная стратегия развития искусственного интеллекта на период до 2030 года согласно указу Президента Российской Федерации от 10.10.2019 г. № 490. Национальной стратегией развития искусственного интеллекта определяются цели и основные задачи развития искусственного интеллекта в России, а также меры,

направленные на его использование в целях обеспечения национальных интересов и реализации стратегических национальных приоритетов, в том числе в области научно-технического развития. Национальной стратегией развития искусственного интеллекта является основой для разработки (корректировки) государственных программ, федеральных и региональных проектов, плановых и программно-целевых документов государственных корпораций, государственных компаний, акционерных обществ с государственным участием, стратегических документов иных организаций в части, касающейся развития искусственного интеллекта.

## **Нанотехнологии в области ИИ**

По мере того, как ученые расширяют возможности машинного обучения, количество ресурсов, требующихся для обучения все более сложных моделей нейронных сетей, стремительно растет. Новая область ИИ, называемая аналоговым глубоким обучением, обещает более быстрые вычисления при меньших затратах энергии. Программируемые резисторы являются ключевыми строительными блоками аналогового глубокого обучения. Повторяя массивы программируемых резисторов в сложных слоях, можно создать сеть аналоговых искусственных нейронов и синапсов, которые выполняют вычисления подобно цифровой нейронной сети. Затем эту сеть можно обучить сложным задачам ИИ, таким как распознавание медицинских изображений и обработка естественного языка у больных.

Междисциплинарная группа исследователей из Массачусетского технологического института решила расширить пределы скорости ранее разработанного ими типа аналогового синапса, созданного человеком. В процессе изготовления они использовали практичный неорганический материал, который позволяет их устройствам работать в 1 000 000 раз быстрее, чем предыдущие версии, что также примерно в 1 млн раз быстрее, чем синапсы в человеческом мозге. Более того, этот неорганический материал также делает резистор чрезвычайно энергоэффективным. Новый материал также совместим с кремниевыми технологиями. Это ноу-хау позволило создавать программируемые резисторы нанометрового масштаба, которые значительно увеличивают скорость обучения нейронной сети, резко снижая энергетические затраты.

Аналоговое глубокое обучение быстрее и энергоэффективнее, чем его цифровой аналог, по двум основным причинам:

- вычисления производятся в памяти, поэтому огромные объемы данных не передаются из памяти в процессор и обратно.
- аналоговые процессоры выполняют операции параллельно. Если размер матрицы увеличивается, аналоговому процессору не требуется больше времени для выполнения новых операций, поскольку все вычисления происходят одновременно.

Ключевой элемент новой технологии известен как протонный

программируемый резистор. Эти резисторы, размеры которых измеряются в нанометрах, расположены в виде массива, как на шахматной доске. В человеческом мозге обучение происходит за счет усиления и ослабления связей между нейронами, называемых синапсами. Глубокие нейронные сети уже давно используют эту стратегию, когда веса сети программируются с помощью алгоритмов обучения. В случае с новым процессором увеличение и уменьшение электрической проводимости протонных резисторов обеспечивает аналоговое машинное обучение.

Идея о том, что оптимизированный диоксид кремния PSG, который представляет собой порошкообразный влагопоглощающий материал, может обладать высокой протонной проводимостью при комнатной температуре без использования воды, оказалась весьма плодотворной. Твердый электролит PSG обеспечивает сверхбыстрое движение протонов. Наносекундный временной масштаб означает, что мы близки к баллистическому или даже квантовому режиму туннелирования протона в экстремальном поле. Поскольку протоны не повреждают материал, резистор может работать миллионы циклов, не выходя из строя.

Теперь, когда продемонстрирована эффективность программируемых резисторов, планируется их крупносерийное производство. Тогда можно будет изучить свойства массивов резисторов и увеличивать их количество до размеров, чтобы из них можно было выстраивать целые системы. Эта работа демонстрирует значительный прорыв в области био подобных резистивных устройств памяти. Эти полностью твердотельные протонные устройства основаны на тонком управлении протонами в атомном масштабе. Они похожи на биологические синапсы, но на порядки быстрее (<https://tenchat.ru/media/2529372-novoye-oborudovaniye-uskoritvychisleniya-dlya-ii>)

В дополнение к вышесказанному хотелось бы привести несколько успешных примеров внедрений популярных технологий искусственного интеллекта в отечественном здравоохранении:

#### DEERMIND HEALTH

Искусственный интеллект обрабатывает всю информацию о пациенте, анализирует все его симптомы и выдаёт список рекомендаций лечащему врачу, который в результате ставит точный и окончательный диагноз.

#### ADA

Искусственный интеллект обрабатывает всю информацию о пациенте, анализирует все его симптомы и даёт консультации напрямую пациенту, подсказывает ему, к какому врачу стоит обратиться и предлагает удалённую консультацию со специалистом.

#### SENSE.LY

Данная программа предназначена для людей, которые недавно прошли длительный срок лечения и у них имеются хронические заболевания. Система выдаёт оповещение о наступлении времени приёма лекарств, необходимости наблюдения у врача, структури-



рует данные о состоянии пациента и отправляет статистику его лечащему врачу.

#### QTrobot

Данный робот предназначен для терапии детей с заболеваниями аутистического спектра. Такие больные с трудом могут контактировать с окружающими, т.к. почти не в состоянии воспринимать чужие эмоции и с трудом выражают свои. Чем старше становится человек, тем труднее ему приходится из-за развития болезни. Поэтому если не уделить должного внимания данной болезни в раннем возрасте, впоследствии справиться с ней трудно. QTrobot предназначен для детей в возрасте от четырёх лет. Он общается с больными с помощью слов, жестов и различными выражениями лица. Такой робот помогает ребёнку со временем научиться распознавать эмоции и настроение окружающих людей, чтобы общаться с ними. В ходе опытной эксплуатации выяснилось, что дети с аутизмом уделяют роботу в среднем в два раза больше внимания, чем лечащему врачу.

## **5. ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ ДЛЯ ЗДРАВООХРАНЕНИЯ ОТ MCKINSEY**

Международная корпорация McKinsey в 2023 году приобрела одну из ведущих мировых компаний по операциям с ИИ и машинным обучением Iguazio, которая в последствии была интегрирована с QuantumBlack. Это подразделение McKinsey занимается экспериментами и инновациями в области ИИ. Iguazio прославилась своими суперпроизводительными платформами ИИ нового поколения, которые предоставляют возможность организовывать и унифицировать процессы дата майнинга между разными командами разработчиков, а также централизованно интегрировать и управлять информационными потоками из множества источников. Платформа искусственного интеллекта Iguazio универсальна и может быть развернута как локально, так и в публичных или гибридных конфигурациях. Она включает расширенные функции контроля происхождения данных, отслеживания и конфиденциальности, необходимые в строго регулируемых отраслях, таких как здравоохранение; также она может использоваться параллельно с языковыми моделями. Для обеспечения ответственного использования ИИ поколения платформа Iguazio имеет встроенные защитные механизмы на каждом этапе развертывания, помогающие предотвращать такие риски, как: утечки данных, предвзятые результаты, галлюцинации ИИ и нарушения прав интеллектуальной собственности. Платформа также поддерживает эффективную структуру управления.

### **Институт здравоохранения McKinsey**

Институт здравоохранения McKinsey / *McKinsey Health Institute* / *MHI* как цифровая платформа с ИИ был создан для помощи ме-

дицинским и социальным работникам, а также педагогам в распознавании у детей признаков ментальной травмы, деэскалации психотравмирующих ситуаций, чтобы действовать более продуктивным и исцеляющим образом. Типичный сценарий, который разыгрывается в начальной школе таков: травмированный ребенок часто капризничает; для поддержания порядка учитель велит ему успокоиться и сидеть тихо. Такие дети всегда реагируют резко негативно, начиная бурно протестовать; обстановка в классе накаляется прямо на глазах. Платформа помогает пользователю глубже понимать суть происходящего, подсказывая, как и какие нужно задавать детям вопросы, например: «Может быть тебе поможет, если ты посидишь за дверью 5-10 минут? Я могу попросить кого-нибудь поговорить с тобой». Это дает ребенку передышку, а следовательно, и свободу действий в фрустрирующей его ситуации, помогая ему постепенно прийти в себя. В итоге инцидент автоматически исчерпывается, и уже не нужно обращаться в администрацию за дисциплинарным решением проблемы. Это всего лишь единичный пример того, что делает развёрнутая на платформе МНН бесплатная онлайн-программа Strong Resilient Youth, разработанная совместно с *Mayerson Center for Safe and Healthy Children at Cincinnati Children's Hospital Medical Center / CCHMC* и помогающая заинтересованным лицам развивать навыки распознавания и реагирования на детские психотравмы.

Концепция данной программы, основанной на 4-ех принципах Управления служб наркотической зависимости и психического здоровья США, заключается в том, что программа является руководством не по лечению, а обучению стейкхолдеров пониманию того, что психотравмы имеют повсеместное распространение и оказывают мощное негативное влияние на окружающий социум. Пользователи совершенствуются в распознавании травм и правильном реагировании при столкновении с неадекватным поведением человека, проявляющего признаки травмирования. В результате ответственные люди воздерживаются от совершения непродуманных действий, которые только усугубляли бы и без того болезненную ситуацию.

Забота о психическом здоровье детей во время пандемии COVID-19 послужила толчком к разработке программы Strong Resilient Youth. Когда мир вошел в режим самоизоляции, а школы перешли в онлайн-режим, медицинская и педагогическая общественность была крайне обеспокоена стрессом в семьях и тем, как это может отразиться на детях в условиях роста случаев домашнего насилия и жестокого обращения с детьми. В Институте здравоохранения McKinsey возникло сильное желание сделать что-то общественно полезное и легкодоступное для помощи малообеспеченным семьям с нездоровыми детьми. Объединенные команды МНН и CCHMC работали со специализированными клиниками и медицинскими ассоциациями со всего мира, занимающимися проблемами психического здоровья и защиты детей, чтобы разработать обучающие датасеты

для платформы ИИ. В подготовке программы также участвовали практикующие специалисты McKinsey Academy, а их коллеги из McKinsey Digital управляли техническими аспектами проекта. В течение года платформа была развернута в лечебных и образовательных учреждениях Среднего Запада США. Учитывая высокую общественную значимость и востребованность программы, в качестве пользователей могли выступать все желающие, начиная от неравнодушных домохозяек, детских спортивных тренеров, сотрудников гуманитарных и благотворительных организаций, и заканчивая профессиональными психоаналитиками и практикующими педиатрами, которые специализируются на помощи наиболее уязвимым слоям населения. Программа была переведена на ряд европейских языков; в США она стала доступна через Национальный совет по психическому благополучию сертифицированным инструкторам по оказанию первой помощи подросткам и молодежи в области психического здоровья.

Деятельность корпорации McKinsey помогла пролить свет на неявные причины детской слепоты. Традиционно обычные люди полагают, даже включая некоторых врачей, что слепота — это последствие заболевания органов зрения. Однако, как оказалось, у многих слепых детей глаза совершенно здоровы. Вместо этого их слепота вызвана корковым нарушением зрения *CVI / Cortical visual impairment*, связанным с повреждением областей обработки зрительных образов в головном мозге. Невозможно выявить какую-то одну причину возникновения CVI. Однако исследования показывают, что CVI возникает при состояниях, влияющих на способность разных частей головного мозга работать скоординированно, утилизировать получаемую информацию и формировать опыт. Опыт каждого человека с CVI уникален, но у всех людей с CVI есть нарушения работы мозга. В эпоху широкого использования методов нейровизуализации и общего интереса к нейронаукам, можно было ожидать, что такие нарушения будут видны при исследовании. Однако только в редких случаях CVI подтверждается наличием крупных структурных дефектов мозга. Большинство случаев — это нарушение способности мозга объединять и обрабатывать информацию. К сожалению, на сегодня невозможно с помощью тех или иных методов сканирования распознать нарушения работы нейронных сетей.

Существует много общеизвестных причин CVI, и по мере развития диагностики выявляется все больше причин и состояний, при которых встречается такое нарушение зрения. Согласно текущим исследованиям причинами CVI могут быть повреждения мозга, связанные с:

- гипоксически-ишемической энцефалопатией новорожденных;
- перивентрикулярной лейкомаляцией. Это повреждение головного мозга, при котором происходит гибель ткани мозга в зонах, прилегающих к боковым желудочкам;

- черепно-мозговыми травмами. Кстати, в случае черепно-мозговых травм CVI может встречаться и у взрослых;
- неонатальной гипогликемией. Это состояние, при котором уровень сахара в крови новорожденного ребенка ниже нормы;
- широким спектром генетических синдромов – синдромов Вильямса, Ретта, CDKL5, трисомии 21, Уэста, Питта Хопкинса, Малана и т.д., описаны также некоторые генетические варианты, собственно, CVI;
- метаболическими нарушениями;
- патологически протекающей многоплодной беременностью;
- пороками развития центральной нервной системы (полимик-рогирией, агенезией мозолистого тела, микроцефалией);
- эпилепсией/судорожными состояниями;
- гидроцефалией;
- перинатальным и детским инсультом;
- осложнениями, связанными с недоношенностью;
- материнской зависимостью от наркотиков и алкоголя.

Согласно данным современных исследований, CVI часто встречается при нарушениях неврологической природы. Это всё вышеперечисленное, а также церебральный паралич, темповые задержки развития, расстройство аутистического спектра, синдром Дауна. CVI может выявляться при генетических синдромах: врожденном нарушении гликозилирования CDG типа 1A, дефиците комплекса II, детской нейроаксональной дистрофии, синдроме Пелицеуса-Мерцбахера, синдроме Мовата-Уилсона и целом ряде генетических аномалий, не связанных с теми или иными синдромами: AHDC1, NGLY1, NR2F1, PGAP1, ACP6, AMOT, ARHGEF10L, ATP6V1A, DCAF6, DLG4, GABRB2, GRIN1, GRIN2B, KCNQ3, KCTD19, RERE, SLC1A1, SLC25A16, SLC35A2, SOX5, UFSP2, UHMK1, ZFP30, ASTN-1.

Поскольку нарушение зрения при CVI имеет причину в мозге, оно чаще всего проявляется у детей с другим неврологическим диагнозом. Оно может проявляться по-разному, например, в виде трудностей с распознаванием материалов, двигательными нарушениями, расстройством ориентации в пространстве. Эти факторы могут затруднить диагностику CVI, оставляя детей без необходимой им специализированной помощи. В международной неправительственной организации для слепых детей *Perkins School for the Blind* в Бостоне заявили, что CVI на самом деле представляет собой настоящую скрытую эпидемию. Им было всегда очень трудно собирать и обобщать информацию о том, насколько широко распространена CVI на самом деле. Всё начало меняться с момента, когда к решению проблемы подключился институт здравоохранения McKinsey, который осуществил масштабное исследование и представил критический анализ полученных данных. Стояла задача понять, в какой степени хроническая венозная недостаточность диагностируется в США. В результате было выявлено порядка 180 тыс. случаев хронической венозной недостаточности, хотя по официальным данным числилось



лишь 24 000. Работа авторитетно подтвердила то, о чем специалисты *Perkins School* подозревали давно: число случаев оказалось гораздо больше того, что они могли себе даже представить.

А ведь сбор информации для анализа распространенности CVI был весьма нелегким делом, так как в базах данных не было ни одного уникального диагностического кода медицинского страхования для идентификации CVI. Команда McKinsey начала с данных, предоставляемых страховыми компаниями, занимающимися сбором информации о пациентах, а затем связалась с клиническими экспертами, чтобы понять, как идентифицировать пациентов с сердечно-сосудистой патологией в ворохе реально анонимных данных. Затем программисты McKinsey воспользовались добытой информацией для построения алгоритма, разработав методику машинного обучения для поиска детей с сердечно-сосудистыми заболеваниями среди общего массива данных о пациентах.

Алгоритм опирался на несколько характеристик из датасета пациентов с диагнозом CVI для оценки по сравнению с пациентами с отсутствием такого диагноза. Используя эти факторы, он научился определить вероятность наличия CVI у пациентов без диагноза. Было обнаружено, что на каждого пациента, которому поставили диагноз, приходилось до 4-ех больных, которым диагноз не поставили. При этом это считалось консервативной оценкой. Команда McKinsey визуализировала данные, наложив на них всю доступную информацию из медицинских организаций и клиник, которые участвовали в проекте, по различным факторам социального риска, включая доступ к врачам, процент незастрахованных пациентов, уровень их благосостояния и транспортную доступность лечебных учреждений. ИИ действительно сгенерировал новый информационный массив по проблеме CVI, что было абсолютно невозможно сделать без его помощи.

Полученные результаты в сочетании с прочей информацией о здоровье населения помогли *Perkins School* целенаправленно скорректировать стратегию оказания медицинской помощи пациентам с заболеваниями сердца и сосудов, применяя самые эффективные методы лечения и подкрепляя свои инициативы данными от McKinsey. Эта работа включает создание учебных материалов для медицинских работников, педагогов, обучающихся лиц с нарушениями зрения и других специалистов, таких как физиотерапевты или логопеды. В обновленные обучающие курсы по диагностике CVI были включены современные способы тестирования в клинических условиях, особенно для глухонемых пациентов. Была развернута работа по пропаганде создания медицинского кодекса для CVI. Выработанные подходы в будущем окажутся особенно полезны в популяциях и районах, в которых фиксируется высокая распространенность CVI, и где ресурсы здравоохранения ограничены. Сегодня врачи со всего мира, занимающиеся проблемами CVI, также как и родители страдающих этим заболеванием детей могут обратиться в *Perkins School* за методической поддержкой и посетить ежегодно проводимую в Бостоне

конференцию по данной теме. Поскольку CVI является неврологическим расстройством, при правильной терапии дети имеют реальные шансы на улучшения функции зрения. «Когда ребенок начинает видеть, это похоже на чудо, но вы не можете совершать такие чудеса без методического руководства, и это то, что мы предоставляем», — так говорят в *Perkins School*, потому что теперь у них есть бесценные данные, которые были предоставлены корпорацией McKinsey.

### **Гибридный искусственный интеллект: сила цифровых технологий + сила людей**

Методология *QuantumBlack AI*, использует мощь ИИ от McKinsey в сочетании с человеческим пониманием. Она сочетает гибкость многопрофильных команд, глубочайшую экспертизу в предметной области с передовым ИИ и силу понимания отрасли здравоохранения McKinsey. Данный баланс раскрывает потенциал гибридного ИИ — конкурентного преимущества, помогающего медицинским организациям по всему миру внедрять инновации и процветать (рис. 12).

Перечислим руководящие принципы ГИИ от McKinsey:

- гармонизация людей и искусственного интеллекта. Методология *QuantumBlack AI* используем возможности технологий и людей для создания роста, который приносит пользу медицинским организациям, одновременно поддерживая процветающее, устойчивое и инклюзивное общество.

- использование отраслевого опыта здравоохранения. Глубокие познания в области машинного обучения и ИИ помогают *QuantumBlack AI* быстрее анализировать сложные данные, определять, где находится наибольшая ценность, стимулировать внедрение и масштабировать решение.

- ускорение воздействия. Рабочие протоколы *QuantumBlack AI*, фирменные инструменты и опытная команда помогают пользователям увидеть ценность наших моделей с первого дня и быстро их развернуть.

- уверенность и доверие. Общее понимание улучшает взаимодействие организаций, людей и технологий. *QuantumBlack AI*, ясно давая понять, откуда берутся идеи проектов, предоставляет уверенность в том, что решения его платформ ИИ являются социально ответственными, высоко моральными и законными.

- разработка моделей ИИ ведётся совместно с пользователями. Разработчики *QuantumBlack AI* являются партнерами по трансформации всех производственных процессов медицинской организации, они всегда работают вместе с будущими пользователями не только над созданием новых платформ, но и над их внедрением, а также над предоставлением персоналу возможностей, необходимых для дальнейшего профессионального роста.

- закрепление достигнутого. *QuantumBlack AI* придерживается принципа устойчивого долговременного сотрудничества с пользователями, постоянно обучая их и закрепляя достигнутые успехи.



**Рис. 12.** Компоненты гибридного искусственного интеллекта (ГИИ).

- сосредоточение на потребностях, а не на технологиях. Каждый проект индивидуален. Технологически-независимый подход *QuantumBlack AI* означает, что разработчики сосредоточены на создании лучшего решения для каждой ситуации и одинаково легко строят модели с чистого листа или работают над совершенствованием существующей инфраструктуры.

- сборные команды профессионалов. Разнообразие ролей и специализация меняют правила игры. Проектирование и создание решений ИИ для обеспечения влияния на бизнес-процессы требуют разнообразия в составе команд, которые объединяют профессиональную экспертизу и глубину технических знаний. К созданию модели ИИ привлекаются люди с нужными навыками и в нужное время.

- инновации – это единственная константа. *QuantumBlack AI* инвестирует в продвижение технологических новшеств и применение передовых знаний, инструментов и методов для решения самых сложных задач, а также активно делится своими достижениями с медицинским сообществом. Пожертвовав инструмент с открытым исходным кодом Kedro, который является основой большинства проектов *QuantumBlack AI*, для Linux Foundation, корпорация McKinsey стремится максимально расширить границы сотрудничества с другими компаниями.

Kedro – это библиотека кода, которую можно использовать для создания конвейеров данных и машинного обучения, строительных блоков любого аналитического проекта. Пользовательская база Kedro продолжают расти, достигнув более чем 200 000 ежемесячных загрузок и более 100 организаций и предприятий, которые выбирают Kedro в качестве своего стандарта для кодирования. Например, команда в NASA использовала Kedro для моделирования схем воздушного движения, а Telkomsel,

крупнейший поставщик беспроводных сетей в Индонезии, использует Kedro в качестве стандарта в своей организации науки о данных. Передача интеллектуальной собственности – это инновационное направление для McKinsey. Но, по словам старших партнеров McKinsey и руководителей QuantumBlack, этот шаг выражает твердое намерение к разработке инновационных продуктов в области ИИ, которые помогают создавать устойчивый и инклюзивный рост социального благополучия населения. Это также представляет революционный сдвиг от эпохи экспериментирования к масштабированию решений ИИ. Теперь будущее Kedro связано с гораздо более широким кругом стейкхолдеров из разных отраслей экономики, географических регионов и технологий, которые приносят разные точки зрения и могут применять Kedro для большего числа вариантов использования. Расширение круга разработчиков – прекрасная идея; теперь каждый желающий может внести свой вклад в разработку Kedro: писать код, формировать стратегии продукта, отслеживать варианты использования и голосовать за решения, которые влияют на проекты. Таким образом, корпорация McKinsey прошла путь от владельца Kedro до пользователя Kedro и тем самым открыла широкие горизонты для внедрения инноваций, благодаря которым будут созданы ранее неизвестные модели ИИ с открытым исходным кодом.

Экосистема стратегических альянсов McKinsey предоставляет клиентам возможности генеративного ИИ / *Gen AI*, который уже проник в нашу жизнь и обещает стать преобразующей технологией ближайшего времени. Мы переживаем стремительный темп инноваций, а это значит, что организации здравоохранения больше не могут полагаться на выбор создания или покупки готовых решений ИИ для внедрения и масштабирования. Чтобы превзойти конкурентов в нынешних условиях, когда *Gen AI* становится все более сложным и обладающим новыми модальностями, такими как видео, аудио, текст, а также все более утонченным технологическим уровнем, медицинским учреждениям необходимо переходить от осторожной проверки надежности концепций внедряемых платформ ИИ к масштабным стратегиям их использования.

Поскольку 90% пилотных проектов не достигают полной производительности, учреждениям здравоохранения нужны соратники, которые помогут им справиться со сложностями принятия и масштабирования ГИИ. Специалистам по внедрению необходимо одновременно справляться с множеством задач: обрабатывать неструктурированные данные, писать передовые алгоритмы, строить правильную ИТ-архитектуру, наращивать возможности, управлять изменениями и экспертизой в предметной области, проч. Мало кто может делать это в одиночку, и именно поэтому сотрудничество с коллегами имеет решающее значение.

Экосистема McKinsey – это решение, которое необходимо всем пользователям продуктов корпорации для внедрения и масштабирования ГИИ. Руководители здравоохранения видят растущую



ценность доступа к ресурсам и передовым технологиям ГИИ. Поэтому экосистема представляет собой централизованный портал для всех стейкхолдеров, позволяющий использовать экспертные знания и решения из 19 стратегических технологических альянсов во всех частях технологического стека: от облачной инфраструктуры до поставщиков LLM *Large Language Models*, т.е. языковых моделей, работающих с большим количеством дополнительных параметров, а также охватывающих такие темы, как машинное обучение, архитектуры данных и наращивание возможностей. Сегодня экосистема открыта и уникальна по своей сути. Чтобы удовлетворить потребности и стремления большинства пользователей, была принята открытая архитектура, созданы стратегические альянсы с ведущими игроками во всех основных технологических стеках.

Благодаря такому подходу удалось объединить самые передовые мировые технологические компании, включая лучших нишевых игроков с активами McKinsey, включая недавние приобретения, такие как Candid для облачных вычислений, Caserta для архитектуры данных и Iguazio для машинного обучения и фундаментальных модельных операций. К тому же экосистема строится на давнем сотрудничестве McKinsey с AWS, Google Cloud, IBM, Microsoft, Salesforce и SAP, добавляя ГИИ в качестве критично важного инновационного компонента совместной работы. С таким широким спектром ведущих поставщиков технологий экосистема может предоставить пользователям лучшие инновации во всех направлениях ГИИ, так как все решения проходят дорожные испытания по всему миру. Тесные отношения, установленные благодаря нашим стратегическим альянсам, и глубокие знания, полученные в результате внедрения их решений, позволяют *QuantumBlack* создавать и разворачивать масштабные сложные приложения искусственного интеллекта промышленным и безопасным способом, включая более 150 крупных проектов искусственного интеллекта за последние месяцы.

## 6. НЕЧЕТКИЕ НЕЙРОННЫЕ СЕТИ В МЕДИЦИНЕ

**Родоначальник современной неклассической логики Н.А. Васильев**

Среди плеяды отечественных учёных, научные заслуги которых получили мировое признание есть мыслители, которые одарили человечество такими идеями, реальная ценность и пионерский смысл которых оказался осознанным лишь десятилетия спустя. Именно к этой плеяде принадлежит профессор кафедры философии Казанского университета Николай Александрович Васильев, в начале XX века предложивший в высшей степени оригинальные логические системы, едва ли не в корне порывавшие с тысячелетними традициями классической логики Аристотеля. Роль Н.А. Васильева в логике в определенном смысле можно сравнить с

ролью Лобачевского в геометрии: идеи Лобачевского положили начало неевклидовой геометрии, а идеи Васильева лежат у истоков неаристотелевой, и в этом плане также неклассической, логики. Васильев начинает изложение своей концепции с констатации того, что уже в логике XIX в. замечается глухая оппозиция против традиционного деления суждений по количеству на общие, частные и единичные, занимавшего в логике Аристотеля исключительно важное положение. Все попытки усовершенствовать это деление выливались лишь в придание ему новой формы. Камень преткновения между тем, согласно Васильеву, лежал в истолковании частных суждений. Ученый показывает, что фактически частные суждения относительно понятий/правил суть общие, а вот суждения относительно вещей, подчиняются обычному делению, и суждения о понятиях и суждения о вещах требуют различной логики. Для суждений о вещах остается справедливым закон исключенного третьего – один из основных законов аристотелевой логики; для суждений же о понятиях необходим закон исключенного четвертого. Поэтому Васильев провозгласил, что закон исключенного третьего должен быть совершенно удален из «скрижали законов мысли».

Васильев предпринял попытку построить новый вид логики, названной им позже воображаемой по аналогии с воображаемой геометрией Лобачевского, путем отказа от одного из основных законов аристотелевой логики: закона *не/противоречия*, всегда принимавшегося за аксиому. Оказалось, что и без этого закона также получаются вполне стройные и замкнутые системы, т.е. аристотелева логика является одной из возможных, равно истинных логик. Таким образом, в дополнение к закону исключенного третьего Васильев выводил также закон не/противоречия из «скрижали законов мысли». Между тем Васильев доказывал, что эти законы обладают глубоким смыслом как эмпирические обобщения и сохраняют свою силу в логике реальных вещей. Если отбросить этот закон, то наряду с утвердительными и отрицательными суждениями становится возможным ввести еще один, отличный от упомянутых, вид суждения, который Васильев назвал индифферентным. Для логики, которая оперировала бы тремя видами суждений, нужен уже не закон исключенного третьего, а закон исключенного четвертого. По мере усложнения устройства воображаемых миров, усложняется и логика, которая может быть не только двумерной как аристотелева, но, вообще говоря, любого количества измерений.

Однако не все логические законы представляют собой эмпирические обобщения в качестве материального аспекта логики. В любой логике имеются законы, делающие возможным само рассуждение – формальный аспект логики. Разграничение формального и материального аспектов в логике предполагает разграничение двух формулировок закона не/противоречия. Одно дело, когда закон не/противоречия запрещает одновременное существование двух

несовместимых признаков предмета, а другое – когда он гласит, что одно и то же суждение не может одновременно быть истинным и ложным. Первое можно отбросить, как это и делается в воображаемой логике, а второе сохраняет силу для любой мыслимой логической системы. Васильев предложил назвать это законом абсолютного разграничения истины и лжи, или законом несамопротиворечия. Минимум логических законов, необходимых для логического рассуждения, составляет металогику как науку о структурах, общих для всех мыслимых логик.

Классическая логика и математика вынуждены ограничивать выразительные возможности своих языков из-за потенциальной опасности внутренних противоречий, парадоксов. Можно сказать, что само стремление построить непременно непротиворечивую систему является своего рода ограничением, обратная сторона которого – феномен недоказуемости свойства непротиворечивости системы ее внутренними свойствами, выраженными теоремами Геделя. А между тем именно стремление доказать формальную непротиворечивость и являлось главным побудительным мотивом ограничительной тенденции. Отсюда вытекает, что непротиворечивость – вовсе не обязательное свойство теоретической системы. Главное, чтобы она была нетривиальной, т.е. не все суждения, сформулированные на ее языке, были равно доказуемыми. Паранепротиворечивая логика и предназначена для изучения противоречивых, но нетривиальных систем. В ней существенно ослабляется связь между доказуемостью и истинностью, а принцип невыводимости из посторонних посылок имеет более фундаментальный характер по отношению к принципу не/противоречия.

Предвидение множественности логических систем, принадлежащее Н.А. Васильеву, сегодня реализовано: построены паранепротиворечивые теории множеств, моделей, алгебраических систем. Появились серьезные и многообещающие исследования в русле паранепротиворечивого подхода к проблемам искусственного интеллекта, экспертных систем и теоретического программирования. А ведь на самом деле трудно себе представить, что еще в начале XX века были высказаны идеи, которые ныне руководят развитием в высшей степени нетрадиционных разделов современной математической логики. К сожалению, Н.А. Васильев не успел перевести свою систему на язык символической логики, а сама система содержит большое разнообразие идей, порой неясно выраженных. Поэтому существуют многообразные формальные модели его взглядов, среди которых наибольший интерес в настоящее время вызывает идейно близкая многозначной логике *нечеткая логика / fuzzy logic*, которая обладает очевидным прикладным эффектом. Она позволяет осуществлять перенос операций с вероятностными законами квантового мира в мир логических рассуждений, давая приближительные истинностные оценки, основанные на теории нечетких множеств (Максимов Д.Ю., 2016) [16].

Fuzzy logic еще называют нечеткозначной логикой. Она является основой рассуждений, в которых используются точно или количественно неопределимые понятия. Нечеткая логика предназначена, в первую очередь, для анализа систем, в которых имеют место человеческие рассуждения и размытые понятия. Если эта логика в настоящее время активно используется для автоматизации сложных технологических процессов, для обеспечения процедур разработки и верификации компьютерных программ, в логических исследованиях естественного языка, в моделировании процессов принятия решений субъектом в той или иной ситуации. Высокая результативность этой логики была доказана в работах по выявлению причин языковых противоречий, логических и семантических антиномий, возникающих в научном познании, а также при решении задач минимизации вреда, который может быть нанесен противоречивой информацией, поступающей в информационно-поисковые системы медицинских организаций. Если в самом начале своего развития паранепротиворечивые логики, к которым относится fuzzy logic, создавались, прежде всего, как средство разрешения и устранения парадоксов в теории множеств, в логике высказываний и логике предикатов, то в дальнейшем они стали эффективно использоваться в естественных, технических, социальных и гуманитарных науках. Так, например, в настоящее время они успешно помогают решению проблем, связанных с необходимостью выражения непротиворечивым образом поступающей в локальную информационную сеть противоречивой информации или несовместимых данных, т. е. в тех ситуациях, когда допускается одновременная истинность или обоснованность некоторого высказывания вместе с его отрицанием. Их аппарат работает в вероятностных и индуктивных рассуждениях, в теории нечетких понятий, в создании нетривиальных онтологических моделей, при необходимости уточнения содержания тех или иных общенаучных понятий, например, «доказательство» и «опровержение», «объяснение» и «предсказание», а также отдельных методологически значимых понятий, таких, как «контрфактическое высказывание», «определимость терминов в теории» и проч. (Яшин Б.Л., 2023) [41].

### **Слияние нечеткой логики и нейронных сетей – апофеоз медицинского ИИ**

Нечеткие нейронные сети / FNN – это уникальная комбинация нечеткой логики и искусственных нейронных сетей, которая породила мощнейший инструмент, способный обрабатывать сложные и неопределенные данные. Поскольку Fuzzy logic есть математическая структура, используемая для представления неопределенности в данных, а искусственные нейронные сети нужны для изучения и прогнозирования на основе моделей данных, то слияние этих двух технологий привело к разработке нечетких нейронных сетей, обрабатывающих большие данные *Bid*



*data*, которые имеют высокую степень неопределенности. Эти сети используются в широком спектре приложений, включая медицинскую робототехнику, обработку изображений/снимков и естественного языка. Вот некоторые термины из области нечетких нейронных сетей:

1. *Нечеткая логика*: направление в математике, которое позволяет описывать и представлять неопределенность в данных. Она основана на принципе, что факты/понятия могут быть частично истинными или частично ложными. В нечеткой логике переменным присваивается определенная степень членства в наборе данных и не присваивается двоичное значение TRUE или FALSE. Например, в системе, которая определяет, является ли человек высоким или нет, переменная «рост» будет представлена определенной степенью членства в датасете ростовых показателей. Она будет иметь значение от 0 до 1, где 0 означает «не высокий», а 1 - «совершенно высокий».

2. *Искусственные нейронные сети*: тип алгоритма машинного обучения, имитирующий функции человеческого мозга. Они состоят из взаимосвязанных узлов, которые обрабатывают информацию и учатся на данных. Такие модели способны обучаться и делать прогнозы на основе изученных данных.

3. *Нечеткие нейронные сети*: сочетая в себе принципы нечеткой логики и искусственных нейронных сетей, представляют собой мощный инструмент, который может обрабатывать данные, имеющие сложную и неопределенную структуру. Эти сети способны обучаться и делать прогнозы на основе такого рода данных.

4. *Применение нечетких нейронных сетей*: нечеткие нейронные сети полезны в широком спектре приложений, например, они могут использоваться для управления движением роботизированного объекта в неопределенной среде. При обработке медицинских изображений эти сети могут использоваться для распознавания патологических образований на изображениях, которые могут быть частично скрытыми или трудно идентифицированными. В обработке естественного языка эти сети помогают понимать смысл выражений и слов, которые имеют несколько значений и, соответственно, могут быть интерпретированы по-разному.

Fuzzy logic с недавнего времени стала довольно популярным подходом для борьбы с неопределенностью и неточностью в сфере здравоохранения. Этот метод как бы имитирует мышление человека, когда истинность информационного сообщения выражается в виде степени, а не бинарным термином истина/ложь. Нечеткая логика обеспечивает гибкий и интуитивно понятный способ моделирования сложных систем, которые трудно описать, используя традиционные математические модели. Вот несколько ключевых направлений использования нечеткой логики:

1. *Системы управления*: нечеткая логика обеспечивает эффективный способ управления сложными и нелинейными системами. Она позволяет создавать правила, которые отражают экспертные

знания и могут адаптироваться к изменениям в системе здравоохранения, что делает ее незаменимым инструментом медицинского менеджмента.

2. *Распознавание шаблона*: нечеткая логика может использоваться для задач распознавания образцов, таких как обработка медицинских изображений, распознавание речи и почерка. Нечеткие классификаторы могут обрабатывать неточные и неоднозначные данные, которые распространены в реальном мире. Нечеткие нейронные сети также могут использоваться для извлечения признаков и уменьшения размерности, что может повысить точность классификации.

3. *Принятие решений*: нечеткая логика может использоваться в задачах принятия решений, таких как оценка риска, оптимизация и медицинская диагностика. Нечеткие модели принятия решений могут обрабатывать неопределенную и неполную информацию и предоставлять лицам, принимающим решения, ряд вариантов на выбор. Например, нечеткая модель принятия решений может быть использована для принятия решения о лучшем плане лечения пациента на основе имеющейся симптоматики, истории болезни и факторов риска.

Слияние нечеткой логики с ИИ имеет несколько преимуществ. Во-первых, это позволяет нейросетевым платформам мгновенно справляться с крайне сложными задачами с наибольшей степенью точности. Во-вторых, это позволяет алгоритмам учиться на своих ошибках, со временем улучшая свою производительность. В-третьих, это позволяет им функционировать в условиях неопределенности и принимать решения на основе неполной или неоднозначной информации. Наконец, это позволяет платформам моделировать сложные системы и делать точные прогнозы. Вот некоторые из определяющих характеристик нечетких нейронных сетей:

1. *Нечеткие наборы данных*: нечеткие нейронные сети используют их для представления входных и выходных значений. Такие наборы допускают более нюансированное представление данных, присваивая степень членства каждому элементу в наборе. Например, нечеткий набор может использоваться для представления роста, где каждый человек имеет определенную степень членства в разных интервалах высоты – низкорослый, среднего роста, высокий.

2. *Нечеткий вывод*: нечеткие нейронные сети используют нечеткий вывод для получения выходных значений в зависимости от введенных параметров. Этот процесс отображения входных значений на выходные осуществляется с использованием нечеткой логики, что включает в себя определение нечетких правил, которые регулируют взаимосвязь между значениями ввода и вывода. Например, если входное значение «высокое кровяное давление», а выходное значение «принять гипотензивный препарат», нечеткое правило может быть определено как «если артериальное давление выше 149/99 мм рт. ст., то есть показание для приема гипотензивных препаратов».

3. *Алгоритмы обучения*: нечеткие нейронные сети используют алгоритмы обучения для оптимизации нечетких правил и параметров. Алгоритмы обучения корректируют веса и смещения сети, чтобы минимизировать ошибку между прогнозируемыми и фактическими выходами. Существуют различные типы алгоритмов обучения, используемых в нечетких нейронных сетях, таких как обратное распространение, генетические алгоритмы и оптимизация роя частиц.

Изначально искусственный интеллект был основан на традиционной модели нейронной сети. С появлением нечетких нейронных сетей FNN произошло объединение нечеткой логики и традиционных нейросетевых моделей с целью создания более надежного и эффективного инструмента машинного обучения. Давайте рассмотрим, как работают нечеткие нейронные сети, включая их архитектуру, процесс обучения и практические приложения.

1. *Архитектура*: нечеткие нейронные сети состоят из 3-ех слоев: входные, нечеткие и выходные слои. Входной уровень получает входные данные, которые затем преобразуются в нечеткие наборы в нечетком слое. Нечеткий слой применяет нечеткую логику к входным данным, что позволяет сети справляться с неопределенностью и неточностью. Выходной слой обрабатывает нечеткие наборы и производит конечный вывод.

2. *Процесс обучения*: учебный процесс нечетких нейронных сетей включает в себя сочетание контролируемого и неконтролируемого обучения. Сеть обучается с использованием набора пар ввода-вывода, что позволяет ей научиться создавать точные выходы для данного входа. При обучении без присмотра модель сама классифицирует данные по различным нечетким наборам, шаг за шагом повышая общую точность выводимой информации.

3. *Практическое применение*: нечеткие нейронные сети имеют широкий спектр практического применения, они особенно хороши в области распознавания и идентификации объектов. Например, нечеткие нейронные сети незаменимы при распознавании рукописных медицинских записей, распознавания смысла смазанной речи больных, для диагностики заболеваний, а также в области медицинского менеджмента.

Нечеткие нейронные сети FNN предлагают уникальный подход к машинному обучению, который помогает преодолеть ограничения традиционных моделей нейронных сетей. Комбинируя нечеткую логику с технологией нейронной сети, разработчики могут создавать более эффективные и точные модели машинного обучения, которые находят широкий спектр практического применения. Нечеткие нейронные сети – это нетривиальный инструмент, который выводит искусственный интеллект на ранее недостижимый уровень. Сочетание нечеткой логики и нейронных сетей стало необычайно эффективным подходом в решении сложных проблем здравоохранения. Это идеальное решение для медицинских приложений, которые ши-

роко используются для распознавания шаблонов и оптимизации процесса принятия решений. Нечеткие нейронные сети имеют целый ряд преимуществ, которые делают их бесценным инструментом в области ИИ:

1. *Повышенная точность вывода*: нечеткие нейронные сети имеют возможность обеспечить более точный вывод, чем традиционные нейронные сети. Они могут решать сложные проблемы, учитывая степень неопределенности и двусмысленности в данных.

2. *Снижение уровня сложности*: нечеткие нейронные сети могут уменьшать сложность системы/проблемы, разбивая ее на более мелкие и более управляемые части. Таким образом, становится проще моделировать самые сложные системы и обрабатывать громадные массивы данных.

3. *Надежность*: нечетким нейронным сетям можно доверить обработку шумных и неполных данных. Они также могут анализировать недостающие данные и принимать решения на основе ограниченной информации.

4. *Адаптивность*: нечеткие нейронные сети могут адаптироваться к новым ситуациям и изменяющейся среде. Они могут учиться на ранее незнакомых форматах данных, со временем улучшая свою производительность.

5. *Интерпретируемость*: нечеткие нейронные сети обеспечивают четкость и краткое увеличение представления данных. Самое главное – они могут объяснить, как алгоритм пришел к тому или иному конкретному решению, предоставив исчерпывающее представление об использованных данных.

Поэтому не случайно, что нечеткие нейронные сети *FNN* успешно используются в диагностике различных заболеваний. Они могут анализировать данные пациента и обеспечить точные прогнозы на основе симптомов и анамнеза пациента. Нечеткие нейронные сети также могут быть использованы в маркетинге медицинских услуг для прогнозирования рыночных тенденций и принятия инвестиционных решений на основе неопределенных данных. Однако, как и любая другая технология, нечеткие нейронные сети имеют свои ограничения. Важно знать о них, чтобы избежать переоценки возможностей этих сетей и выработки соответствующих решений для их преодоления. Вот те ограничения, которые следовало бы здесь упомянуть:

- *Требования к данным*: нечеткие нейронные сети требуют больших объемов данных для точного обучения. Во многих реальных приложениях получение больших объемов данных может быть сложным, особенно когда данные являются дорогостоящими, трудоемкими или трудно собираемыми. Это может привести к неточному или неполному обучению, что может повлиять на производительность платформы ИИ.

- *Отсутствие интерпретации*: вывод нечеткой нейронной сети порой трудно интерпретировать, особенно когда проблема сложна. В отличие от обычных нечетких систем на основе правил, нечеткие



нейронные сети зачастую не предоставляют четких формулировок своих решений. Это может быть проблемой в приложениях, где интерпретируемость является критической, такой как медицинская диагностика или финансовое прогнозирование.

- *Ограниченная гибкость*: нечеткие нейронные сети не так гибки, как обычные нейронные сети. Нечеткий логический компонент сети накладывает некое ограничение, делающее поведение модели ригидным в новых ситуациях. Это может быть проблемой для медицинских приложений, работающих в динамичной и постоянно меняющейся среде.

- *Вычислительная сложность*: вычислительная сложность нечетких нейронных сетей может быть высокой, особенно при работе с большими наборами данных. Неизбежным следствием этого являются длительный период обучения и замедленная производительность на малых и средних серверных мощностях. Естественно, это совсем не способствует масштабируемости платформ и делает их малоприспособленными для функционирования в реальном времени.

- *Трудности при настройке*: настройка параметров нечеткой нейронной сети может быть непростой задачей. В отличие от обычных нейронных сетей, которые имеют четко определенный набор легко настраиваемых параметров, нечеткие нейронные сети имеют целые ансамбли параметров, которые довольно трудно оптимизируются. Это может вести к неоптимальной производительности, требовать значительного количества рабочего времени программистов и мощных вычислительных ресурсов.

Блок нечеткого интерфейса предоставляет входной вектор многослойной нейронной сети в качестве ответа на лингвистические утверждения. Впоследствии нейронная сеть обучается генерировать требуемые выводы / решения. Порядок работы механизма нечеткого вывода самостоятельно определяется многослойной нейронной сетью (рис. 13).

Нейронная сеть предлагает нечеткой модели вычислительные характеристики обучения, получая взамен интерпретацию и ясность представления о функциональной архитектуре платформы. Нечеткие нейронные сети *FNN* также называемые нейро-нечеткими системами *NFS* используют получаемые от искусственной нейросети методы аппроксимации, к которым относятся нечеткие множества и нечеткие правила, для определения параметров нечеткой системы *NFS* (рис. 14).

Разберёмся в типах нейро-нечетких систем, начиная с кооперативной *NFS*. В моделях кооперативных *NFS* искусственная нейронная сеть и нечеткая система работают независимо. При этом процесс получения параметров искусственной нейронной сетью из нечеткой системы может протекать как в автономном режиме, так и онлайн (рис. 15, 16).

В верхнем левом квадрате рисунка 16 представлен автономный режим работы *NFS*, когда нечеткие правила, определяемые на основе обучающих данных в сочетании с нечеткими множествами/набо-

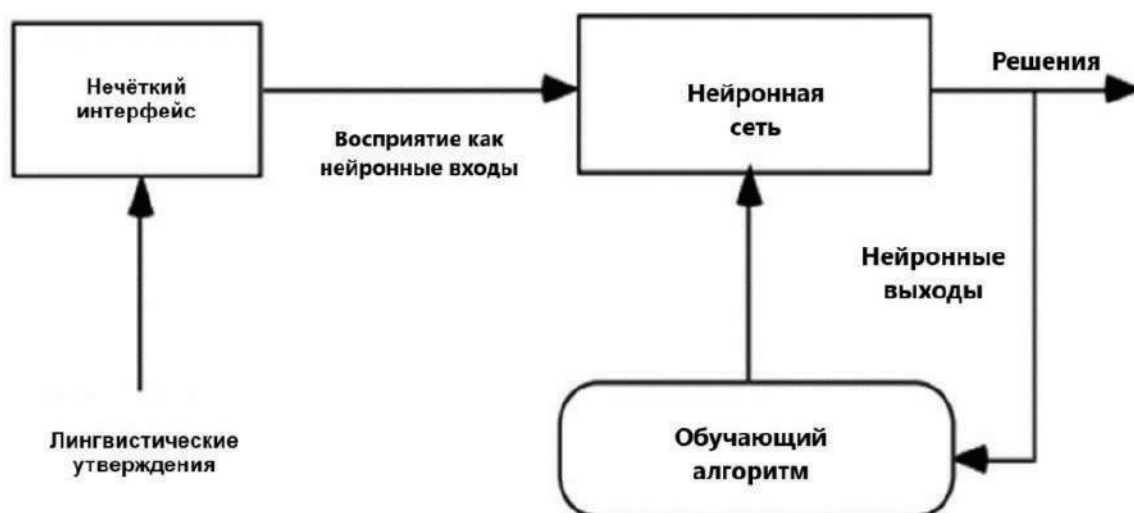


Рис. 13. Блок нечеткого интерфейса FNN.



Рис. 14. Модель нечеткого нейрона.

рами, используются для формирования выводов нечеткой системы. В правом верхнем квадрате нечеткая нейронная сеть изучает нечеткие множества/наборы из заданных обучающих данных. Тут также осуществляется автономное определение вывода нечеткой системы. В левом нижнем квадрате рисунка изображена модель работы, когда нечеткие правила и функции принадлежности должны быть определены заранее, чтобы нечеткая система могла изучить все параметры функции принадлежности онлайн. Для повышения эффективности шага обучения необходимо измерять ошибку, поэтому в правой нижней модели вес правила, который интерпретируется как влияние правила, определяется для всех нечетких правил нейронной сетью. Таким образом осуществляется как онлайн, так и офлайн определение вывода. Кооперативная NFS использует искусственные нейронные сети только на начальном этапе. На основе обучающих данных



Рис. 15. Механизм работы кооперативной нейро-нечеткой системы.

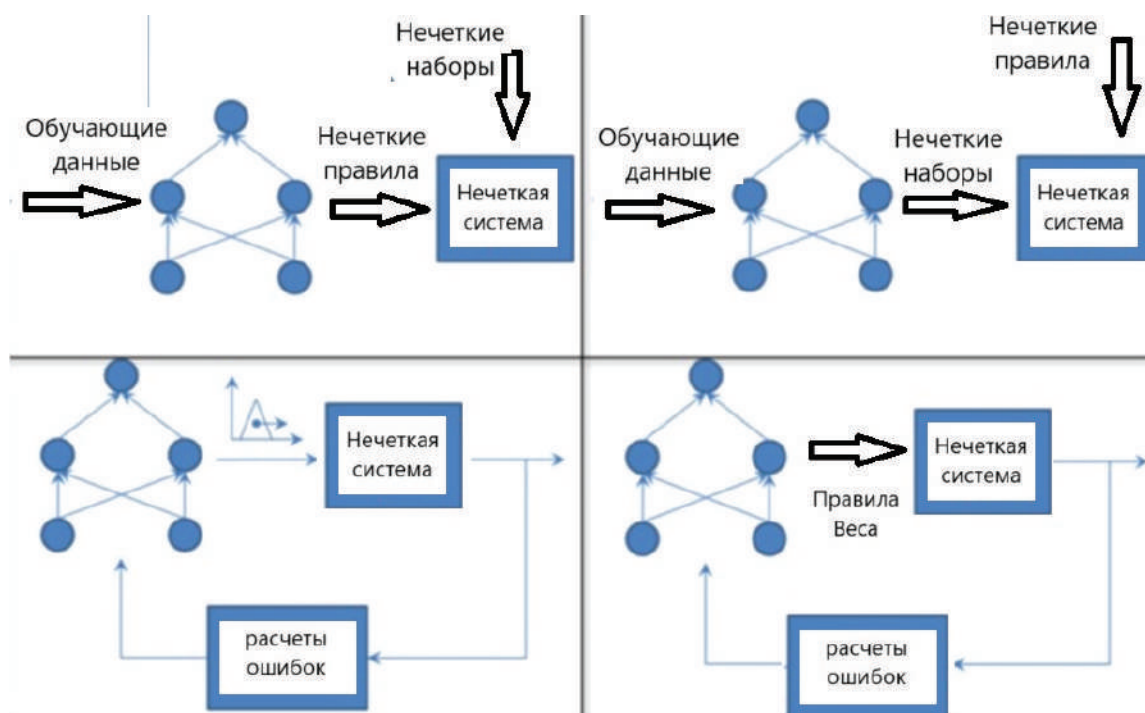


Рис. 16. Различные типы кооперативных нечетких нейронных сетей.

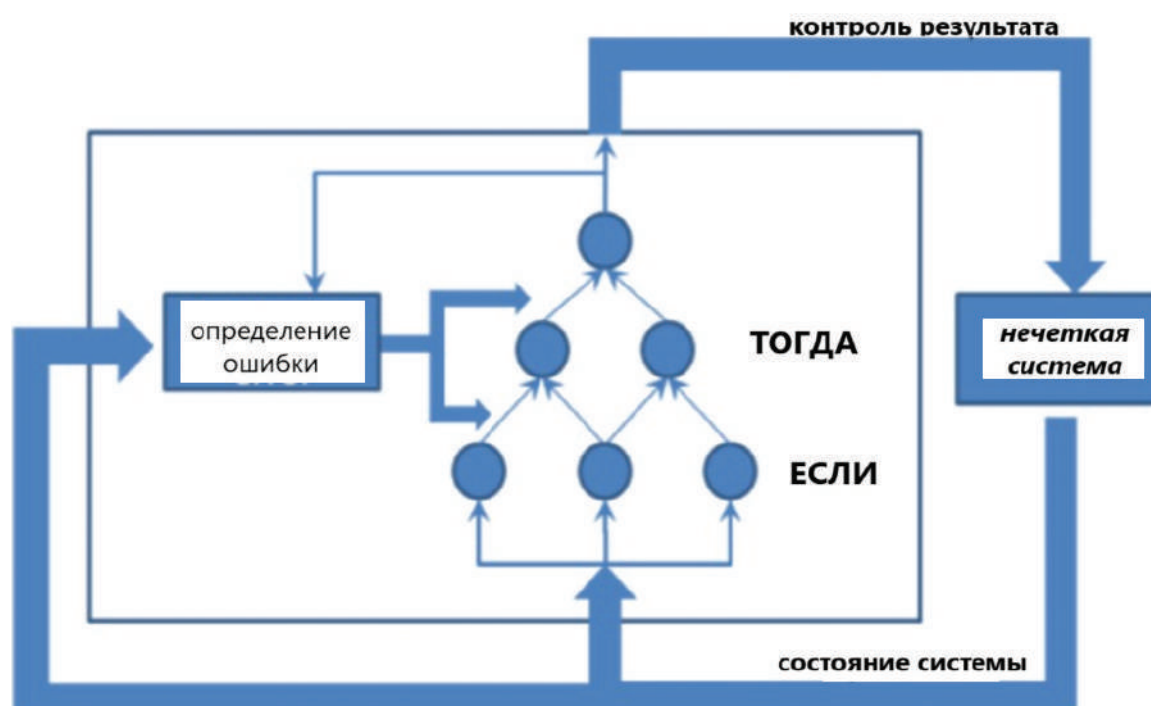
эти сети устанавливают своего рода подблоки нечеткой системы, последующее удаление которых приводит к реализации уже самой по себе нечеткой системы как функциональной платформы.

В *параллельной нейро-нечеткой системе*, нейронная сеть и нечеткая система постоянно функционируют совместно, причем нейронная сеть предварительно обрабатывает входные данные нечеткой системы (рис. 17).

Гибридные нейро-нечеткие системы используют нейронные сети для расчета определенных параметров нечеткой системы. В этом случае архитектура *гибридной NFS* дает большое преимущество, поскольку нечеткая система и нейронная сеть не должны взаимодействовать друг с другом. Кроме того, эти системы могут обучаться как в онлайн режиме, так и в режиме офлайн (рис. 18).



**Рис. 17.** Параллельная нейро-нечеткая система.



**Рис. 18.** Гибридная нейро-нечеткая система.

Благодаря своей высокой эффективности методы нечеткой логики заняли ведущее положение в области медицинского искусственного интеллекта, потому что они помогают быстро проводить диагностику и выдавать точные прогнозы течения заболеваний. Благодаря нечеткой логике в 2019 г. был создан нечеткий классификатор артериальной гипертензии. При этом оказалось, что системы нечеткого вывода вышеописанных типов представляют собой наилучшие архитектурные решения для создания указанной классификации. Нечеткая логика также применялась для оценки риска гипертонии. Ещё в 2018 г. была разработана модель, которая объединила нейронные сети и нечеткую логику для этой цели. Нечеткие системы были ключевой частью этого исследования, поскольку они обеспечивали регулируемость уровня классификационной неопределенности.



Наилучшие результаты с превосходной производительностью относительно поставленной задачи были продемонстрированы именно гибридной моделью *NFS*.

В настоящее время гибридные системы являются мощным инструментом для решения сложных проблем, поскольку два или более методов мягких вычислений могут использоваться одновременно для решения конкретной проблемы, и тем самым уменьшая ее сложность. Кроме того, гибридные системы направлены на повышение эффективности и мощности рассуждений, а также выразительности изолированных интеллектуальных систем. Мягкие вычисления являются противоположностью жестких вычислений, поскольку они терпимы к неточности, неопределенности, частичной правде и приближению. Для разработки гибридной моделью *NFS* использовались методы мягких вычислений, такие как искусственные нейронные сети и нечеткая логика, поскольку эти методологии имеют ряд общих характеристик, включая тот факт, что они являются свободными оценщиками моделей, которые можно настраивать или обучать для улучшения их производительности.

Искусственная нейронная сеть – это система обработки информации, которая имеет определенные характеристики производительности, общие с биологическими нейронами. Искусственные нейронные сети были разработаны как обобщения математических моделей человеческого познания или нейронной биологии. Модульный принцип компоновки нейронных сетей основан на стратегии «разделяй и властвуй», которая заключается в разложении задачи на менее сложные подзадачи, каждая из которых изучается экспертами отдельно, и только потом повторно происходит обучение каждого модуля для решения всей проблемы в целом. Для объединения результатов обучения модулей используются интеграторы ответов, такие как усреднение, победитель получает все, нечеткая логика, вентильная сеть и проч.

Нечеткие системы нужны в качестве интеграторов ответов, поскольку таким образом удастся справляться с неопределенностью в решении, поскольку, как уже указывалось ранее, нечеткая логика – это тип логики, который включает приблизительные, а не точные способы рассуждения. Её можно рассматривать как попытку построить модель человеческого рассуждения, которая отражает его приблизительный и качественный характер. Конечная цель состоит в том, чтобы предоставить основу для приблизительного рассуждения с использованием неточных предложений, основанных на теории нечетких множеств, аналогично классическому рассуждению с использованием точных предложений, основанных на классической теории множеств (Melin P., et al., 2018) [113].

Нечеткие системы также активно применяются для диагностики болезни Паркинсона. Нечеткая нейронная система позволяет эффективно выделять в обследуемых контингентах больных именно тех людей, которые страдают паркинсонизмом. Системы *NFS* совершенно незаменимы при проверке методик классификации круп-

ных массивов медицинских данных путем построения моделей нечеткого вывода или нечетких экспертных платформ. Анализ больших объемов информации, связанных с болезнью Паркинсона, дал мощный толчок научному направлению, когда системы на основе знаний в сочетании с методами интеллектуального анализа и нечеткими алгоритмами принятия решений, а также классификаторами искусственных нейронных сетей служат полезными инструментами сопоставления клинической информации с наборами оцифрованных данных путем использования наборов нечетких правил. В отношении болезни Паркинсона адаптивная нейро-нечеткая экспертная система продемонстрировала более высокие показатели точности, чем обычная компьютерная экспертная система. Кроме того, адаптивная нейро-нечеткая экспертная система показала более высокий уровень чувствительности и специфичности. Приведем методологию построения адаптивной нейро-нечеткой экспертной системы Adaptive Neuro-fuzzy expert system:

Этап 1. Формирование датасета на основе клинической информации, полученной из медицинских учреждений.

Этап 2. Пошаговый запуск адаптивно-нейронной системы нечеткого вывода *adaptive-Neuro fuzzy inference system / ANFIS* по команде:

а) запуск системы в работу; б) загрузка данных из депозитария; в) начало тренировки датасета гибридным методом; г) установка периодичности обучения набора данных; д) повторные тестирования и проверки; е) формирование 5-ти слойной структуры ANFIS; ж) 1-ый слой используется для принятия нечетких функций принадлежности; з) 2-ой слой состоит из фиксированных узловых компонентов, отражающих весомость представляемых ими правил. В этом слое по команде нормирующего оператора выполняется операция дополнения. Выход слоя 2 является произведением входящих сигналов; и) в 3-ем слое ANFIS, также содержащим фиксированные узлы, происходит нормализация правил, поступающих из предыдущего слоя. Они известны под названием нормализованных сил срабатывания; л) В 4-ом слое каждый узел уже является адаптивным узлом, где нормализованная сила срабатывания из предыдущего слоя 3 потенцирует наборы последующих параметров; м) 5-ый слой ANFIS состоит из фиксированных узлов, помеченных как сумма. Там происходит вычисление общего выхода, исходя из всех входящих сигналов по сформированному алгоритму до тех пор, пока все входные данные не будут классифицированы правильно.

Этап 3. На основе параметров, выбранных из обученных датасетов, производится оценка точности, чувствительности и специфичности созданной адаптивной нейро-нечеткой экспертной системы.

Сельское малограмотное население, проживающее вдоль реки Нил в Египте, имеет самую высокую в мире заболеваемость болезнью Паркинсона. Вторая по величине в мире распространенность этой

болезни обнаруживается среди общин амишей в штатах Пенсильвания и Огайо. Именно по результатам обследования данных сообществ был сделан вывод о том, что точность платформ ИИ со встроенной адаптивной нейро-нечеткой экспертной системой намного выше, чем при использовании экспертных систем традиционных конфигураций (Kaur P., et al., 2017) [92].

### **Проблематика эффекта нелинейности нейронных сетей в представлении выпускников цифровых кафедр медицинских вузов**

Цифровизация здравоохранения – это общемировой тренд. Перед будущим поколением российских медицинских работников во весь рост встаёт непростая задача приобретения профессиональных компетенций в области электронного здравоохранения. В скором времени молодые врачи должны будут уметь общаться с разработчиками программного обеспечения на равных. Без знания секретов функционирования медицинских нейронных сетей выпускнику медицинского вуза будет весьма сложно стать высоко эффективным специалистом в избранной сфере деятельности. Не случайно в 2022 году в 114 университетах страны, являвшихся участниками программы «Приоритет 2030», появились первые цифровые кафедры. Проект «Цифровые кафедры» реализуется в рамках федерального проекта «Развитие кадрового потенциала ИТ-отрасли» национального проекта «Цифровая экономика Российской Федерации». Целью проекта «Цифровая кафедра» является тиражирование цифровых компетенций для современного специалиста в области здравоохранения. Студенты изучат алгоритмизации, программирование для решения задач на языке программирования Python, новейшие JS-фреймворки типа React или Angular и системную аналитику. Например, РНИМУ им. Н.И. Пирогова успешно защитил образовательную программу «Основы программирования и бизнес-аналитики для медицинских работников». Программа создавалась инжиниринговым центром РНИМУ совместно с ведущими компаниями ИТ-отрасли, в том числе из Кластера биологических и медицинских технологий Инновационного центра «Сколково». В работе над ней также активно участвовала кафедра медицинской кибернетики и информатики медико-биологического факультета РНИМУ. Для студентов СЗГМУ им. И.И. Мечникова сегодня доступна профессиональная переподготовка по специальности «Интеллектуальные информационные технологии в медицине», обучение по которой уже проходят более 800 желающих. В Сеченовском Университете подготовка ведётся сразу по нескольким направлениям: разработчик цифровых медицинских сервисов, специалист по анализу медицинских данных, разработчик решений виртуальной и дополненной реальности в медицине, DevOps в медицине. Для улучшения профессиональных навыков и знаний для студентов также ведётся набор на дополнительные программы по-

вышения квалификации: цифровая трансформация, цифровая медицина, основы DevOps. Запись на курсы доступна для студентов, проходящих учебу очно или очно-заочно на 2-4 курсе бакалавриата или 3-5 курсе специалитета, а также в магистратуре и ординатуре.

Молодое поколение врачей прекрасно видит, что цифровые технологии начинают играть всё более значимую роль в повышении качества и доступности медицинской помощи как в городах, так и в удаленных местах проживания населения. Медицинские нейронные сети ускоряют и облегчают процессы исследований, помогают совершать новые открытия и снижать риски на этапах клинических испытаний. Применение геномной инженерии, клеточных технологий или роботизированной техники для диагностики и лечения сложных заболеваний, фармацевтическое производство – все это сегодня реализуется с помощью цифровых технологий. Программному обеспечению на основе ИИ не сегодня-завтра предстоит стать основным двигателем роста производительности труда и конкурентоспособности медицинских организаций, ускоряющего темпы развития здравоохранения страны в целом. Поэтому хорошо овладевшие цифровыми компетенциями выпускники в перспективе смогут легче трудоустраиваться, приоритетно участвовать в инновационных проектах, добиваться успехов в научной деятельности и, пользуясь уважением коллег и пациентов, получать стабильно высокий доход.

Вместе с тем, быстрое распространение нейросетевых технологий имеет ряд негативных аспектов, среди которых присутствует глубокая озабоченность студентов-медиков в отношении собственного будущего: они опасаются, что их может заменить ИИ. Такая озабоченность вряд ли может способствовать адекватному восприятию преимуществ искусственного интеллекта студентами медицинских вузов. Известны исследования, указывающие на то, что нейросеть может сократить образовательные возможности для развития способностей клинического суждения и накопления практических навыков врачами-стажерами (Banerjee M., et al., 2021) [49]. Поскольку приложения ИИ изменят роли практикующих врачей и компьютерных средств в медицинском обслуживании, преподаватели и иные заинтересованные лица должны следить за процессом освоения студентами цифровых кафедр ключевых модулей учебной программы, чтобы выявлять и вовремя устранять возможные пробелы в знаниях. Standing Committee of European Doctors (CPME) в 2020 г. постулировал 3 основные группы цифровых компетенций: общие, технические и связанные с отношениями врач-пациент. В идеале выпускникам необходимо наработать достаточно высокий уровень упомянутых компетенций для того, чтобы приступить к их практическому применению. При этом критически важно целеустремленно использовать освоенные навыки, внося максимальный вклад в развитие платформ ИИ на благо пациентов. Неуклонное стремление и энтузиазм при узнавании о новых информационных технологиях в медицине всегда



зависит от множества взаимосвязано варьирующихся факторов, из переплетения которых очень трудно выделить не только главное, но и нечто общее (Roda S., 2021) [139].

В данном контексте для получения релевантной информации относительно уровня технической грамотности молодежной аудитории интересно иметь представление о глубине понимания студентами и выпускниками медицинских вузов какого-либо конкретного вопроса, касающегося особенностей функционирования платформ ИИ, например, проблематики нелинейных эффектов в нейросетях. Нелинейность играет важную роль в построении и функционировании нейронных сетей. Сети, состоящие из линейных элементов, могут выполнять только линейные операции и ограничены в своих возможностях. Нелинейные функции позволяют обеспечить более широкий диапазон задач, которые может решать медицинская нейронная сеть. Линейные функции ограничены в том смысле, что они не могут моделировать сложные зависимости между входными и выходными данными. Следствием этого является то, что линейные сети неспособны эффективно распознавать и классифицировать сложные объекты. В противоположность им, сети с нелинейными функциями активации способны обучаться сложным аспектам данных, показывать более точные результаты, моделируя такие неочевидные взаимосвязи, как нелинейные границы классов или сложные временные зависимости. Кроме того, наличие эффекта нелинейности также позволяет нейронной сети выполнять высокоуровневые задачи, такие как выявление аномалий, выделение важных признаков и создание новых признаков путем комбинирования существующих. Это дает таким нейронным сетям огромное преимущество в анализе больших объемов данных и решении максимально сложных задач, когда высокая точность и гибкость особенно востребованы.

Нелинейность достигается с помощью функций активации, которые применяются к выходу каждого нейрона в нейронной сети. Эти функции нелинейны по своей природе и позволяют сети передавать и обрабатывать информацию в формате разнообразных кривых. Существует большое количество функций активации, имеющих свои особенности и применяемых в различных целях. Наиболее популярными функциями активации являются сигмоид, *ReLU Rectified Linear Unit* и гиперболический тангенс. Выбор функции активации зависит от конкретной задачи и может существенно влиять на качество работы нейронной сети. Вот несколько преимуществ использования нелинейных функций активации в нейронных сетях.

- Способность к аппроксимации сложных функций. Нелинейные функции активации позволяют нейронной сети моделировать сложные и нелинейные зависимости между входами и выходами. Это делает их эффективным инструментом для решения разнообразных задач, таких как классификация, регрессия и генерация контента.

- Избегание проблемы симметрии. Нелинейные функции активации помогают предотвратить проблему симметрии, которая

может возникнуть при использовании линейных функций. Это позволяет сети изучать различные аспекты данных и улучшает обобщающую способность сети.

- **Повышение устойчивости к шуму.** Нелинейные функции активации способны справляться с шумом и артефактами в данных. Они позволяют сети проводить нелинейные операции и учитывать непредсказуемые факторы, что делает их более устойчивыми к внешним помехам.

- **Решение задачи градиентного затухания.** Нелинейные функции активации помогают решить проблему градиентного затухания, которая возникает при обратном распространении ошибки в глубоких нейронных сетях. Они позволяют восстанавливать информацию о градиентах и предотвращать их затухание.

Вышеперечисленные преимущества делают эффект нелинейности главным условием построения качественных медицинских нейронных сетей, без которого невозможно достичь никакой сложности и добиться высоких результатов машинного и глубокого обучения. Без него у нас была бы просто одноуровневая сеть, когда речь могла идти лишь о машинном обучении, а вовсе не о глубоком обучении. Только при глубоком обучении мы имеем возможность фиксировать разделение между каждым слоем нейросети, чтобы выявлять их предназначение. Нелинейное преобразование изменяет пространственное представление данных, давая возможность их исследования с разных точек зрения, а, следовательно, лучшего их понимания. Кроме того, в целях получения качественных прогнозов нейронные сети при помощи механизма прямого распространения обеспечивают пользователя релевантной информацией, извлекаемой ими из огромного массива данных. Таким образом, глубоко обученная нейросеть всегда нацелена на обработку данных для извлечения представлений, наиболее полезных для решения поставленной задачи. Глубокое обучение – это очень мощная технология, именно поэтому фирмы Google, Amazon, Meta и OpenAI используют её для создания искусственного интеллекта завтрашнего дня.

Недавно компания Google выпустила новую библиотеку AutoKeras, благодаря которой больше нет необходимости выполнять Feature Engineering, то есть очистку данных. AutoKeras позволяет создавать модели глубокого обучения без необходимости самостоятельно реализовывать архитектуру. AutoKeras способна автоматически заботиться о выборе структуры слоев, количестве нейронов и других гиперпараметров, например, таких как функции оптимизации и потерь. Теперь провести глубокое обучение нейросети можно буквально с помощью трех строк кода и получить замечательные результаты за считанное время. К тому же в дополнение к выбору подходящей модели для решения поставленной задачи AutoKeras может самостоятельно выполнить предварительную обработку данных, будь то цифры, текст либо изображения. Пользователю достаточно просто выбрать тип задачи и осуществить подготовку к вводу данных, используя лишь 2 строки кода.

Несмотря на то, что AutoKeras действительно позволяет быстро получать ожидаемые результаты, крайне важно хорошо разбираться в том, что вообще происходит внутри системы, знать, как она действует и понимать внутренний механизм работы библиотеки. Проблема состоит в том, что AutoKeras имеет свойство не показывать скрытые алгоритмы глубокого обучения. Также ни в коем случае не стоит забывать о феномене *переобучения*. Это явление возникает тогда, когда нейросеть становится настолько специализированной для определенных данных, что оказывается непригодной для анализа иных реальных данных. Выходом из этой ситуации может служить обучение нейросетевой модели на сугубо обобщенных данных, что, однако же, не всегда приемлемо.

Если вернуться к роли, выполняемой в моделях глубокого обучения функцией активации, то нужно сказать, что каждый нейрон имеет функцию активации и вес. Однако, в то время как весовые коэффициенты назначаются случайным образом всем нейронам, функция активации выбирается вручную инженером машинного обучения ML Engineer. Модель может состоять из тысяч и тысяч нейронов, поэтому ML Engineer физически неспособен назначить функцию активации каждому отдельному нейрону. Выход состоит в том, что функция активации назначается инженером машинного обучения при добавлении нового слоя в модель. Затем функция активации автоматически распределяется между нейронами в этом слое. Таким образом, нейроны в одном и том же слое получают одну и ту же функцию активации. Практически это реализуется с помощью того же AutoKeras, когда функция активации устанавливается при добавлении очередного слоя нейросети. Например, указать функцию активации DENSE сигмовидной функции с 32 нейронами в слое можно следующей командой:

```
model.add(layers.Dense(32, activation='sigmoid'))
```

Кроме того, важно подчеркнуть, что функции активации – это не просто атрибуты уровней глубокого обучения. Помимо алгоритмов глубокого обучения, они также являются функциями, которые можно использовать для изменения данных. В контексте глубокого обучения существует несколько функций активации и важно знать, как их правильно выбрать. Действительно, если неверно выбрать функцию активации, то можно легко саботировать корректную работу всей нейросетевой платформы.

В процессе глубокого обучения модели нейронной сети может обнаружиться, что предопределенные функции активации в библиотеках Python не соответствуют целевой задаче. Допустим, происходит так, что доступные функции активации не дают желаемого результата или есть потребность в применении дополнительных нелинейных преобразований. В этом случае мы можем создать свои собственные функции активации, следуя двум правилам:

1. Функция активации должна быть нелинейной.

2. Функция активации принимает тензор в качестве входных данных.

Здесь стоит сразу оговориться, что классические библиотеки Python предоставляют функции, которые можно применять к вещественным числам. Например, в `math.exp(x)` `x` является вещественным числом. Но чтобы создать функцию активации, нужно использовать функции, применимые к тензорам. Например, в `tensorflow.math.exp(x)` `x` является тензором. С помощью AutoKeras функция активации создается командой:

```
from keras import backend as K
def maFonction(x, beta=1.0):
    return x * K.sigmoid(beta * x)
```

Созданная функция активации может быть добавлена в модель независимо от использования слоя, как в приводимом ниже коде, а затем применяться к выходу каждого нейрона предыдущего слоя (Haifeng J., et al., 2023):

```
from keras.layers.core import Activation
model.add(layers.Dense(16))
model.add(Activation(maFonction))
```

В представлении молодых медиков об ИИ центральное место занимает понятие многослойного перцептрона MLP, с которого, по сути, они начинают освоение нейронных сетей и принципов их обучения. Это одна из элементарных архитектур прямого распространения, состоящая как минимум из 3-х слоёв: входного, скрытого и выходного. Здесь нужно подчеркнуть, что за исключением входных, все нейроны использует именно **нелинейную функцию** активации. Это в очередной раз напоминает нам о важности затронутой нами темы нелинейных эффектов систем ИИ для начинающих профессиональную карьеру врачей. Исторически изобретение перцептрона связано идеей обработки информации живым нейроном. Нейрон получает входящий импульс через свои дендриты, которые передают принятый сигнал телу клетки. Точно так же перцептрон получает входные сигналы из примеров обучающих данных, которые предварительно взвесили и объединили в линейное уравнение, называемое активацией. В чистом виде *MLP* для решения практических задач в области машинного обучения уже давно не используется, но применяется в составе сложных медицинских нейросетей (рис.19).

Платформы с искусственным интеллектом быстро совершенствуются с ускорением вычислений, развитием аппаратного обеспечения и появлением технологий глубокого обучения нейронных сетей. Молодежная аудитория врачей согласна с тем, что ИИ исключительно полезен современной медицине и клинической практике. Ряд приложений на основе ИИ в таких областях как



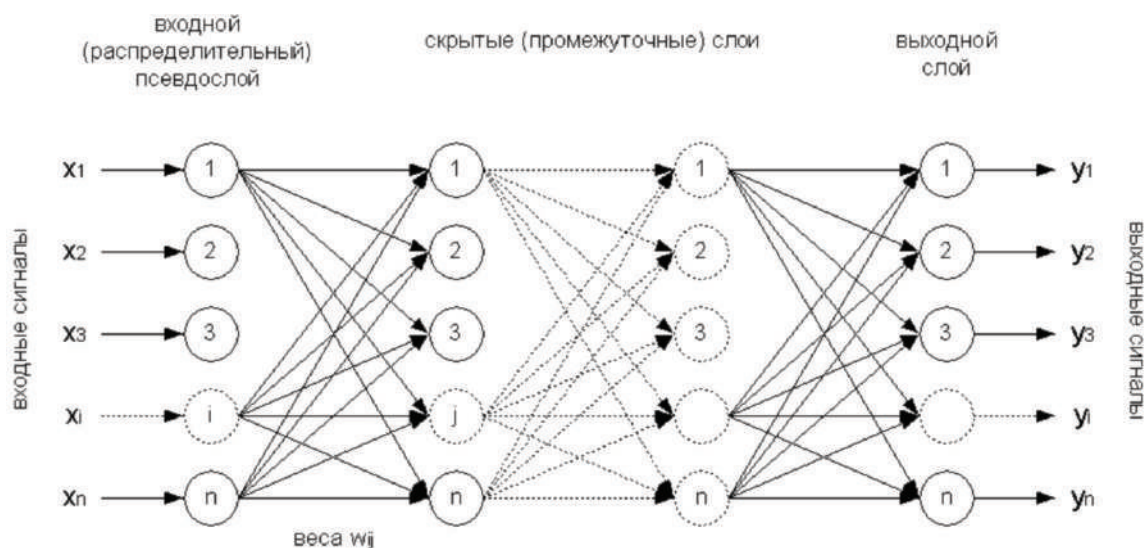


Рис. 19. Структура многослойного перцептрона *MLP*.

клиническая и геномная диагностика, лечение хронических заболеваний, медицинский блокчейн, хирургические роботы глубоко изменили медицинскую практику XXI века. Очевидно, что индивидуальная цифровая грамотность в недалеком будущем будет являться неотъемлемой частью врачебной профессии, но в текущее время именно она лимитирует скорость прогресса в здравоохранении по всему миру. Руководители медицинских организаций единодушны во мнении, что отрасли требуется гораздо большее, чем просто высококвалифицированные медицинские специалисты. Развитие цифровых компетенций персонала является одним из приоритетных и универсальных направлений образования в рамках всех медицинских специальностей. Для качественного внедрения нейросетевых технологий в медицинскую практику необходимо целенаправленное и полноценное обучение медицинских работников на цифровых кафедрах или курсах. Медицинская информатика должна стать фундаментальным навыком, позволяющим каждому выпускнику образовательного учреждения использовать платформы с ИИ для улучшения качества и доступности медицинской помощи, обеспечения полноценного партнёрства врачей и пациентов [84].

Глубокое понимание эффекта нелинейности нейронных сетей является отличным индикатором, свидетельствующим о твёрдой поведенческой установке молодых врачей осваивать базовые принципы функционирования медицинского ИИ. Между прочим, это ещё и косвенное свидетельство того, что сосредоточенность исключительно на освоении клинических дисциплин отвлекает внимание студентов медицинских вузов от изучения технологий ИИ. Для формирования устойчивого ментального намерения овладения цифровыми навыками у студентов и практикующих врачей заинтересованные структуры должны ускоренными темпами внедрять в планы образовательных учреждений учебные про-

граммы по самым современным медицинским нейросетевым платформам, подчёркивая их чрезвычайную актуальность для будущей карьеры выпускника. С этой целью преподаватели медицинского вуза должны создать такую учебную среду, в которой нейросетевые модели перестанут быть отвлеченными от реальной жизни технологическими изысками, о которых пишут в научных журналах, а будут играть роль ключевых инструментов в повседневной клинической практике.

## **ЗАКЛЮЧЕНИЕ**

Давайте подведём некоторые итоги всего, что было сказано касательно будущего медицинского искусственного интеллекта, ещё раз выделив узловые аспекты в этом связанные. Итак:

ИИ уже спасает жизни больных и может спасти еще больше. Прогностические расчёты Европейского альянса за доступ к безопасным лекарствам на 2023 г. показали, что в Европе 163 тыс. человек должны были бы умереть из-за врачебных ошибок, что оказалось довольно близко к реальным причинам смертности. Фактически до 30% врачебных ошибок происходят из-за неэффективной коммуникации врач-больной. ИИ идеально подходит для налаживания системы межличностной коммуникации, предоставляя врачу необходимую информацию в нужное время в правильном контексте, что способствует предотвращению ошибочных решений, спасению жизни пациентов и улучшению результатов лечения. Это лишь одно из следствий способностей алгоритмов ИИ обрабатывать огромные массивы клинических данных, например, визуальную информацию, истории болезни и прочее для помощи в диагностике и оптимизации лечения пациентов. При безопасном и правильном использовании это может экспоненциально расширить масштабы применения достижений доказательной медицины.

ИИ определенно даст возможность медицинским работникам уделять больше внимания непосредственному обслуживанию пациентов. ИИ поможет врачам заниматься своими прямыми профессиональными обязанностями, выделяя больше времени на общение с пациентами, и не тратить избыточных усилий на заполнение медицинской документации и административную работу. До 36% производственных процессов в сфере здравоохранения и социальной помощи можно автоматизировать с помощью ИИ. Такой рост производительности сократит прогнозируемый дефицит в 3,5 миллиона медицинских работников, которые потребуются странам Организации европейского экономического сотрудничества к 2030 году. ИИ также может помочь поставщикам медицинских услуг интегрировать базовые знания и добывать данные о показателях здоровья населения, чтобы исходя из них, предотвращать избыточную смертность больных из-за ограниченных возможностей учреждений здравоохранения и продвигать передовые клинические прак-

тики. Таким образом, ИИ безусловно способен позитивно повлиять на качество работы персонала, качество человеческого взаимодействия и качество результатов.

ИИ, вероятно, сможет защитить цифровую инфраструктуру здравоохранения от угроз, обеспечив надёжную кибербезопасность баз медицинских данных. Сегодня мы наблюдаем постоянный рост числа кибератак на структуры здравоохранения, которые, по прогнозам американских аналитиков, к 2025 году приведут к финансовым потерям в размере до 10,5 триллионов долларов США. Хакеры все чаще используют ИИ для поиска малейших уязвимостей корпоративных информационных сетей. Организации здравоохранения могли бы извлечь уроки из использования ИИ в других отраслях, чтобы вовремя обнаруживать и предотвращать угрозы, а также быстро устранять последствия DOS-атак и мошеннических действий киберпреступников. Стандартизация в области ИИ будет способствовать укреплению доверия пользователей к цифровым инструментам.

ИИ может подсказать руководителям отрасли здравоохранения как извлечь выгоду из такого ценного актива, которым являются статистические данные о здоровье населения и которые в настоящее время крайне плохо используются для принятия решений. По мнению специалистов Корнельского университета, проектирование, разработка и внедрение платформ ИИ в здравоохранении может сильно выиграть от своевременного доступа к качественным данным при соблюдении надлежащих мер защиты. Уже накоплен богатый опыт использования моделей ИИ для обнаружения ранних признаков COVID-19, ускорения разработки вакцин, которые спасли миллионы жизней во всем мире. ИИ будет использоваться для совершенствования методов лечения редких и орфанных заболеваний, научных открытий при выявлении ранее неизвестных заболеваний, новых способов профилактики хронических заболеваний и продвижения персонализированной медицины. Инновационные приложения ИИ будут выглядеть в глазах пользователей более репрезентативными и эффективными, если сотрудничество между системами здравоохранения разных стран начнет развиваться более активно.

Очевидно, что возможности медицинского ИИ значительны, они впечатляют своими перспективами, но требуют решительных действий для устранения существующих барьеров и угроз. Поступательное выстраивание политики управления данными и материально-технического базиса, на которых ответственный и безопасный ИИ может быть разработан, внедрен и масштабирован, принесет пользу всем слоям общества. Это должно быть сделано таким образом, чтобы минимизировать потенциальные риски от нарушений принципов конфиденциальности, безопасности, непредвзятости и прозрачности, но неизменно обеспечивая положительные результаты для всех стейкхолдеров. В связи с этим необходимы проактивные инициативы в разработке стратегий внедрения наиболее надежных архитектурных решений

нейросетевых платформ в целях соблюдения важнейших прав пациентов на конфиденциальность, безопасность и отсутствие дискриминации.

Готовность здравоохранения к изменениям устоявшегося уклада в первую очередь зависит от готовности заимствовать лучшие цифровые решения у других отраслей. Например, технологии, которые используют для рендеринга изображений и отображения теней в компьютерных играх, уже успешно применяют в медицине. Трансформация медицины из искусства и ремесла в индустриальное производство наблюдается в мире всё отчетливее, и это вовсе не замена специалиста на робота с ИИ, а обогащение медицины инновационными цифровыми технологиями, например, из ритейла, авиаперевозок, индустрии туризма и гостеприимства.

Применение ИИ способно произвести революцию в системе здравоохранения за счёт развития таких направлений, как персонализированная медицина, диагностика, разработка новых лекарственных препаратов, робот-ассистированная хирургия, телемониторинг хронических заболеваний, дистанционная помощь пациентам, поддержка принятия правильных медицинских решений, выявление медицинских ошибок. Правда роботов нельзя научить человеческой мудрости врача-эксперта, чувству эмпатии и заботе о больных, взаимопониманию и поддержке своих коллег. Это именно то, что и составляет суть врачебной профессии, и без чего цифровые платформы ИИ окажутся просто бесполезными дорогими игрушками. Таким образом, самая большая проблема для ИИ в здравоохранении заключается не в том, окажутся ли эти технологии достаточно полезными, а в том, чтобы обеспечить их внедрение в повседневную клиническую практику. Со временем клиницисты смогут перейти к задачам, требующим уникальных человеческих навыков, тем, что подразумевают высочайший уровень когнитивных функций. И, пожалуй, единственными проигравшими в этой погоне за ИИ могут стать те медицинские организации, которые откажутся по-настоящему его использовать.

## ОПРЕДЕЛЕНИЯ И ТЕРМИНЫ

**Artificial intelligence (AI)** Искусственный интеллект (ИИ). Исторически сложилось, что под термином ИИ понимается способность нейронной сети имитировать отдельные проявления человеческого интеллекта или даже превосходить его в решении какой-то частной задачи, например, в прогнозировании или в рассуждениях на заданную тему. Однако при написании данной монографии авторы учитывали приоритет машинного обучения, как доминирующей в настоящее время области медицинского ИИ.

**Machine learning (ML)** *Машинное обучение (МО)*. Это область ИИ, которая касается методов, которые обучают нейронную сеть на основе существующих данных выполнять поставленные задачи, например, задачи прогнозирования или классификации.



**Big data (BD) Большие данные.** Термин используется в случаях, когда выборки данных слишком велики, чтобы их можно было адекватно проанализировать с помощью традиционных методов ИИ. В этом случае можно использовать новые методы, такие как глубокие нейронные сети (также известные как «глубокое обучение»).

**Neural networks Нейронные сети (НС).** Искусственные НС представляют собой схемы, состоящие из определенного количества взаимосвязанных нейронов, организованных иерархически в слоях и способных обучаться выполнению очень сложных задач на основе данных. Каждый нейрон действует как своего рода специализированный процессор, который преобразует входные данные в выходные сигналы. Эти преобразования зависят от приложения и извлекаются из доступных данных, специфичных для приложения. Постепенно нейроны объединяют свои выходные данные, слой за слоем, аппроксимируя обработку сложной интегральной функции, пока сеть не выдаст окончательный результат, например диагноз или прогноз заболевания.

**Deep learning (DL) Глубокое обучение (ГО).** DL относится к нейронным сетям с более чем тремя слоями; в этом случае наличие BD необходимо для оценки оптимальных значений параметров для этого более крупного и сложного типа глубокой нейронной сети. Однако не все инструменты искусственного интеллекта и машинного обучения основаны на глубоком обучении или нейронных сетях. Другие методы, такие как деревья решений или машины опорных векторов, не менее широко используются, особенно когда выборка данных недостаточно велика для построения нейронных сетей или глубоких нейронных сетей.

**Модель искусственного интеллекта, алгоритм ИИ, инструмент ИИ.** В специализированной технической литературе под алгоритмом ИИ понимается процедура, используемая для построения модели ИИ для конкретного приложения, следовательно, модель ИИ – это результат работы алгоритма машинного обучения. Другими словами, один и тот же алгоритм ИИ можно использовать для построения моделей (например, прогностических моделей) для множества различных приложений, но модель ИИ специфична для конкретного приложения (например, прогнозирования реакции пациента на определенный вид лечения). Вместе с тем, термины «алгоритмы искусственного интеллекта» и «модели искусственного интеллекта» (или «алгоритмы машинного обучения» и «модели машинного обучения») часто используются как синонимы. Инструменты ИИ – это модели ИИ, которые упакованы для использования конечными пользователями, поэтому они содержат больше, чем просто модель ИИ, например пользовательские интерфейсы. В неспециализированной литературе термины: модели, алгоритмы, инструменты, решения и программное обеспечение ИИ используются взаимозаменяемо, особенно в медицинских кругах.

**Обучение, валидация и данные тестирования.** Обучающие данные – это наборы данных, которые используются разработчиками ИИ для обучения своих моделей ИИ. Данные проверки или валидации также используются разработчиками ИИ. Однако они используются для оптимизации параметров моделей ИИ, чтобы их можно было применять к новым данным, отличным от данных обучения. Другими словами, валидация используется для тонкой настройки моделей ИИ, чтобы сделать некоторые обобщения. Данные тестирования – это уже совсем новые данные, отличные от тех, которые используются для обучения и оптимизации моделей ИИ. Они нужны для оценки работы моделей ИИ; и лучше внешними независимыми оценщиками.

**Medical AI Медицинский ИИ.** Вид искусственного интеллекта, ориентированный на конкретные приложения в медицине и здравоохранении.

**Проектирование, разработка, оценка и внедрение ИИ.** Это некие основные этапы жизненного цикла медицинского ИИ. Сначала разрабатываются инструменты ИИ, как правило, в рамках совместного творчества и сотрудничества между разработчиками ИИ и клиническими экспертами в какой-то конкретной области (а иногда также с привлечением пациентов и нескольких групп разных экспертов, в том числе и руководителей здравоохранения высокого ранга). Разработчики ИИ пишут код для создания и оптимизации моделей ИИ на основе данных обучения и проверки, имеющихся в их распоряжении. Впоследствии модель ИИ оценивается с использованием данных тестирования, отличных от данных обучения и проверки. Инструмент ИИ (модель ИИ с пользовательским интерфейсом) также оценивается конечными пользователями: врачами



**Рис. 20.** Связь между искусственным интеллектом, машинным обучением и глубоким обучением.

и/или пациентами. Если оценка оказалась успешна и убедительна для соответствующих стейкхолдеров (пациентов, врачей, руководителей здравоохранения, регулирующих органов), инструмент ИИ проверяется, утверждается и затем применяется на практике. Описанный конвейер, конечно, является идеальным сценарием, и на практике, безусловно, могут встретиться различные варианты этапов разработки и жизненного цикла ИИ.



## СПИСОК ЛИТЕРАТУРНЫХ ИСТОЧНИКОВ

1. *Абаева О.П., Рябова Е.Н.* Особенности модели взаимоотношения пациентов, перенесших пересадку органов, и врачей центра трансплантации // Национальное здоровье. 2019. №1. С. 23-26. EDN ULAZDU.
2. *Анохин К.В.* Мыслящая гиперсеть // В мире науки. 2021. №5-6. <http://scientificrussia.ru/articles/myslyashchaya-giperset-v-mire-nauki-5-6-2021>
3. *Бердутин В.А., Абаева О.П., Романова Т.Е., Романов С.В.* Применение искусственного интеллекта в медицине: достижения и перспективы. Обзор литературы. Ч.1. // Социология медицины. 2022. Т.21. №1. С. 83-96. doi: 10.17816/socm106054
4. *Бердутин В.А., Абаева О.П., Романова Т.Е., Романов С.В.* Применение искусственного интеллекта в медицине: достижения и перспективы. Обзор литературы. Ч. 2. Социология медицины. 2022. Т.21. №2. С 203-209. doi:10.17816/socm107908
5. *Бердутин В.А., Романов С.В., Запорожцев А.В., Абаева О.П., Романова Т.Е.* Системный подход в медицинских организациях. Учебно-методическое пособие для руководителей медицинских организаций. М.: ФГБУ ГНЦ ФМБЦ им. А.И. Бурназяна ФМБА России, 2021. 140 с. ISBN: 978-5-6046269-6-2
6. *Бердутин В.А., Бердutiна Э.В., Романов С.В., Романова Т.Е., Абаева О.П.* Типологические основы медицинского менеджмента: Учебное пособие. М.: ФГБУ ГНЦ ФМБЦ им. А.И. Бурназяна ФМБА России, 2023. 240 с. ISBN: 978-5-93064-240-7
7. *Бердутин В.А.* Робастное управление человеко-центрированной медицинской организацией // Главврач. 2020. №5. С. 17-34. doi:10.33920/med-03-2005-03
8. *Бердутин В.А., Бердutiна Э.В.* Логистика прикладных решений для бережливого здравоохранения и соционическая типология. Beau Bassin: LAMBERT Academic Publishing, 2020. 206 с. ISBN: 978-620-0-54874-0
9. *Бердутин В.А., Романова Т.Е.* Методология контент-анализа информационного поля цифровой медицины // Главврач. 2023. №4. С. 5-10. doi:10.33920/med-03-2304-01
10. *Бердутин В.А., Романова Т.Е., Романов С.В., Абаева О.П., Дзюбак С.А.* Понятийный аппарат, связанный с цифровизацией здравоохранения: Учебно-методическое пособие. М.: ФМБЦ им. А.И. Бурназяна ФМБА России, 2024. 68 с.
11. *Булычева Е.В.* Искусственный интеллект как новое явление в развитии здравоохранения и медицинского образования (обзор литературы) // Медицинское образование и профессиональное развитие. 2022. Т.13. №3. С. 76-84. doi: 10.33029/2220-8453-2022-13-3-76-84
12. *Власов Я.В., Фомина Т.А., Синеок Е.В., Светозарский С.Н.* Оценка доступности медицинской помощи при заболеваниях сетчатки в субъектах Российской Федерации: мнения врачей и пациентов // Саратовский научно-медицинский журнал. 2024. Т.20. №2. С. 198-202. doi: 10.15275/ssmj2002198.
13. *Гарбук С.В., Шалаев А.Р.* Перспективная структура национальных стандартов в области искусственного интеллекта // Стандарты и качество. 2021. №10. С. 26-33.
14. *Глизница П.В., Тахчиди Х.П., Светозарский С.Н., Бурсов А.И., Шустерзон К.А.* Машинное обучение в диагностике и лечении офтальмологических заболеваний. Head and Neck/Голова и шея // Российское издание. Журнал Общероссийской общественной организации Федерация специалистов по лечению заболеваний головы и шеи. 2022. Т.10. №1. С. 83-90. doi: 10.25792/HN.2022.10.1.83-90.
15. *Итинсон К.С.* Искусственный интеллект как перспективная технология в области медицинского образования и медицины // Карельский научный журнал. 2020. Т.9. №2. С. 16-18.
16. *Максимов Д.Ю.* Логика Н.А. Васильева и многозначные логики // Логические исследования. 2016. Т.22. №1. С. 82-107.



17. *Мамина Р.И., Пирайнен Е.В.* Эмоциональный искусственный интеллект как инструмент взаимодействия человека и машины // ДИСКУРС. 2023. Т.9. №2. С. 35-51. doi: 10.32603/2412-8562-2023-9-2-35-51
18. *Павлов С.В., Абаева О.П.* Социология медицины: тенденции и векторы развития // Социология медицины. 2019. Т.18. №2. С. 128. EDN WDMBBB.
19. Перспективная программа стандартизации по приоритетному направлению «Искусственный интеллект» на период 2021-2024 годы, утверждённая Министерством экономического развития Российской Федерации и Федеральным агентством по техническому регулированию и метрологии 22.12.2020. Электронный ресурс: [https://www.economy.gov.ru/material/news/v\\_rossii\\_poyavyatsya\\_standarty\\_v\\_oblasti\\_iskusstvennogo\\_intellekta.html](https://www.economy.gov.ru/material/news/v_rossii_poyavyatsya_standarty_v_oblasti_iskusstvennogo_intellekta.html)
20. О создании технического комитета по стандартизации "Искусственный интеллект": Приказ Федерального агентства по техническому регулированию и метрологии (Росстандарт) от 25.07.2019 № 1732. Росстандарт. Электронный ресурс: <https://www.rst.gov.ru/portal/gost/home/activity/documents/orders#/order/104460>
21. Об утверждении Концепции развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники на период до 2024 г.: Распоряжение Правительства РФ от 19 августа 2020 г. № 2129-р.
22. *Резаев А.В., Трегубова Н.Д.* Искусственный интеллект и искусственная социальность: новые явления и проблемы для развития медицинских наук // Эпистемология и философия науки. 2019. Т.56. №4. С. 183-199
23. *Решетников А.В.* Социальный институт медицины. Ч.1 // Социология медицины. 2018. Т.17. №1. С. 4-11. doi: 10.18821/1728-2810-2018-17-1-4-11
24. *Решетников А.В.* Социальный институт медицины. Ч.2 // Социология медицины. 2018. Т.17. №1. С. 68-79. doi: 10.18821/1728-2810-2018-17-2-68-79
25. *Романова Т.Е., Абаева О.П., Присяжная Н.В., Романов С.В., Бердугин В.А.* Взаимоотношения врачей и пациентов инфекционного госпиталя в периоды 1 и 2-й волн пандемии COVID-19 // Терапевтический архив. 2024. Т.96. №1. С. 42-44. doi: 10.26442/00403660.2024.01.202562
26. РОБОПРАВО. Исследовательский центр проблем регулирования робототехники и искусственного интеллекта. Электронный ресурс: [http://robopravo.ru/modiel-naia\\_konvientsiia](http://robopravo.ru/modiel-naia_konvientsiia).
27. *Романов С.В., Дзюбак С.А., Романова Т.Е., Бердугин В.А., Абаева О.П.* Опыт внедрения информационной системы с технологиями искусственного интеллекта в работу поликлиники для оптимизации организации периодических медицинских осмотров // Менеджер здравоохранения. 2024. №9. С. 60-66. doi: 10.21045/1811-0185-2024-9-60-66.
28. *Романова Т.Е., Абаева О.П., Романов С.В.* Отношение врачей и пациентов центральной районной больницы к применению технологий цифровой медицины (по результатам медико-социологического исследования) // Саратовский научно-медицинский журнал. 2023. Т.19. №2. С. 180-183. doi: 10.15275/ssmj1902180.
29. *Романова Т.Е., Абаева О.П., Романов С.В., Родина А.А.* Отношение пациентов многопрофильного стационара к процессам цифровизации в современном здравоохранении: одномоментное сплошное исследование // Социология медицины. 2022. Т.21. №2. С. 177-181. doi: 10.17816/socm111928
30. *Романова Т.Е., Родина А.А., Романов С.В., Абаева О.П.* Оценка качества подготовки студентов медицинских вузов в период пандемии COVID-19 для будущей работы в практическом здравоохранении // ОРГЗДРАВ: новости, мнения, обучения. Вестник ВШОУЗ. 2022. Т.8. №1. С. 75-81. doi: 10.33029/2411-8621-2022-8-1-75-81
31. *Романова Т.Е., Абаева О.П., Романов С.В., Родина А.А.* Оценка готовности врача многопрофильного стационара к применению цифровых технологий в практической деятельности // ОРГЗДРАВ: новости, мнения, обучения. Вестник ВШОУЗ. 2022. Т.8. №3. С. 105-111. doi: 10.33029/2411-8621-2022-8-3-105-111

32. Ситнова А.В., Валитов Э.Р., Светозарский С.Н. Применение алгоритмов глубокого машинного обучения на основе многослойной нейронной сети YOLOv8 для идентификации грибкового кератита // Современные технологии в медицине. 2024. Т.16. №4. С. 5-14. doi: 10.17691/stm2024.16.4.01.
33. Танишин Е.С., Танишина Е.Н. Искусственный интеллект в медицинском образовании, преимущества и перспективы // Виртуальные технологии в медицине. 2023. №3. С.17. doi: 10.46594/2687-0037\_2023\_3\_1701
34. О развитии искусственного интеллекта в Российской Федерации: Указ Президента Российской Федерации от 10.10.2019 № 490.
35. О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте Российской Федерации - городе федерального значения Москве и внесении изменений в статьи 6 и 10 Федерального закона "О персональных данных": Федеральный закон № 123-ФЗ от 24 апреля 2020 г.
36. Хазов М.В., Романов С.В., Абаева О.П., Мурыгина М.М. Мотивация врачебного персонала многопрофильной больницы к работе сверх нормативов одной ставки // Проблемы социальной гигиены, здравоохранения и истории медицины. 2015. Т.23. № 2. С. 35-37.
37. Хохлов Ю.Е. Стандарты работы с данными для искусственного интеллекта: ландшафт стандартизации искусственного интеллекта // Информационное общество. 2023. №3. С. 78-96. doi:10.52605/16059921\_2023\_03\_78
38. Черная И.П., Просалова В.С., Николаева А.А. Сквозные технологии как цифровые инновации в здравоохранении и медицинском образовании. Теория и практика общественного развития. 2022. №3. С. 64-73. doi:10.24158/tipor.2022.3.9
39. Чубов С.А. Теоретические и технологические аспекты формирования профессиональной компетентности будущих фармацевтов на основе применения средств искусственного интеллекта // Известия Волгоградского государственного педагогического университета. 2022. Т.2. №165. С. 117-121.
40. Щерабаков М.Г. Эмоциональный интеллект как элемент системы искусственного интеллекта: этические и правовые вопросы // Ученые записки Крымского федерального университета имени В. И. Вернадского. Юридические науки. 2022. Т.8. №3. С. 201-209.
41. Яшин Б.Л. Неклассические логики в современной науке // Философская мысль. 2023. 1; 15-25. doi: 10.25136/2409-8728.2023.1.39350
42. Adedinsewo D., Carter R.E., Attia Z., Johnson P., Kashou A.H., Dugan J.L., et al. Artificial Intelligence-Enabled ECG Algorithm to Identify Patients with Left Ventricular Systolic Dysfunction Presenting to the Emergency Department with Dyspnea. *Circ Arrhythmia Electrophysiol.* 2020;13:8.
43. Adhikari L., Ozrazgat-Baslanti T., Ruppert M., Madushani R.W.M.A., Paliwal S., Hashemighouchani H., et al. Improved Predictive Models for Acute Kidney Injury with IDEA: Intraoperative Data Embedded Analytics. *PLoS One.* 2019;14:4.
44. Ahn J., Connell A., Simonetto D., Hughes C., Shah V.H. Application of Artificial Intelligence for the Diagnosis and Treatment of Liver Diseases. *Hepatology.* 2021;73;6:2546-2563.
45. Almufareh M., Imran M., Khan A., Humayun M., Asim M. Automated Brain Tumor Segmentation and Classification in MRI using YOLO-based Deep Learning. *IEEE Access.* 2024;12:2-20. doi: 10.1109/ACCESS.2024.3359418. URL: [https://www.researchgate.net/publication/377792098\\_Automated\\_Brain\\_Tumor\\_Segmentation\\_and\\_Classification\\_in\\_MRI\\_using\\_YOLO-based\\_Deep\\_Learning](https://www.researchgate.net/publication/377792098_Automated_Brain_Tumor_Segmentation_and_Classification_in_MRI_using_YOLO-based_Deep_Learning)
46. Artificial Intelligence in Healthcare. Applications, Risks, and Ethical and Societal Impacts. European Parliamentary Research Service. Scientific Foresight Unit STOA Panel for the Future of Science and Technology. Brussels, European Union, 2022. 69 p. ISBN: 978-92-846-9456-3. doi:10.2861/568473

47. *Avuçlu E.* Determining the Most Accurate Machine Learning Algorithms for Medical Diagnosis Using the Monk' Problems Database and Statistical Measurements. *Journal of Experimental & Theoretical Artificial Intelligence.* 2023;21:357-376. doi: 10.1080/0952813X.2023.2196984
48. *Azzi S., Gagnon S., Ramirez A., Richards G.* Healthcare Applications of Artificial Intelligence and Analytics: A Review and Proposed Framework. *Appl. Sci.* 2020;10:6553. doi: 10.3390/app10186553.
49. *Banerjee M., Chiew D., Patel K.T., Johns I., Chappell D., Linton N., Cole G.D., Francis D.P., Szram J., Ross J., et al.* The Impact of Artificial Intelligence on Clinical Education: Perceptions of Postgraduate Trainee Doctors in London (UK) and Recommendations for Trainers. *BMC Med. Educ.* 2021; 21:429. doi: 10.1186/s12909-021-02870-x
50. *Beede E., Baylor E., Hersch F., et al.* A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* Honolulu, 2020 Apr 25-30; 1-12. doi: 10.1145/3313831.3376718
51. *Berdutin V.A., Romanova T.E., Romanov S.V., Abaeva O.P.* Problematic Aspects of Medical Artificial Intelligence. *P.1. Sociology of Medicine.* 2023;22; 2:202-211. doi:10.17816/socm619132
52. *Berdutin V.A.* Socionic Vision on Bioethics and Deontology. Beau Bassin, LAP LAMBERT Academic Publishing, 2018. 394 p. ISBN: 978-613-4-97887-3
53. *Boniolo F., Dorigatti E., Ohnmacht A.J., Saur D., Schubert B., Menden M.P.* Artificial Intelligence in Early drug Discovery Enabling Precision Medicine. *Expert Opin Drug Discov.* 2021;1-17.
54. *Camm A.J., Kirchhof P., Lip G.Y., et al.* Guidelines for the Management of Atrial Fibrillation: the Task Force for the Management of Atrial Fibrillation of the European Society of Cardiology (ESC) European Heart Rhythm Association; European Association for Cardio-Thoracic Surgery. *Eur Heart J* 2010;31:2369-2429.
55. *Byambasuren O., Sanders S., Beller E., Glasziou P.* Prescribable mHealth apps Identified from an Overview of Systematic Reviews. *NPI Digital Medicine.* 2018;1;1: Art. 12.
56. *Campello V., et al.* Multi-Centre, Multi-Vendor and Multi-Disease Cardiac Segmentation: The M&Ms Challenge. *Medical Image Computing and Computer Assisted Intervention.* 2021 Dec;40;12:3543-3554.
57. Carnegie Mellon University. Atlantic Causal Inference Conference – 2018: Data Challenge. 2018. URL: <https://www.cmu.edu/acic2018/data-challenge/index.html>
58. *Challen R., Denny J., Pitt M., et al.* Artificial Intelligence, Bias and Clinical Safety. *BMJ Qual Saf.* 2019;28;3:231-237.
59. *Christodoulou E., Ma J., Collins G.S., Steyerberg E.W., Verbakel J.Y., Van Calster B.* A Systematic Review Shows no Performance Benefit of Machine Learning Over Logistic Regression for Clinical Prediction Models. *Journal of Clinical Epidemiology.* 2019;110:12-22.
60. *Clay H., Stern R.* Making Time in General Practice. Primary Care Foundation. 2015:1-83.
61. *Daley K.* Two Arguments against Human-Friendly AI. *AI and Ethics,* 2021;1;4:435-444. doi:10.1007/s43681-021-00051-6
62. *Daneshjou R., Vodrahalli K., Novoa R.A., et al.* Disparities in Dermatology AI Performance on a Diverse. Curated Clinical Image Set. Cornell University. 2022:1-13. <https://arxiv.org/ftp/arxiv/papers/2203/2203.08807.pdf>
63. *Dash S., Shakyawar S.K., Sharma M., Kaushik S.* Big Data in Healthcare: Management, Analysis and Future Prospects. *Journal of Big Data.* 2019;6:54. doi: 10.1186/s40537-019-0217-0
64. *Davis S.E., Lasko T.A., Chen G., Siew E.D., Matheny M.E.* Calibration Drift in Regression and Machine Learning Models for Acute Kidney Injury. *Journal of the American Medical Informatics Association.* 2017;24;6:1052-1061.



65. Dawoodbhoy F.M., Delaney J., Cecula P., Yu J., Peacock I., Tan J., Cox B. AI in Patient flow: Applications of Artificial Intelligence to Improve Patient Flow in NHS Acute Mental Health Inpatient Units. *Heliyon*. 2021;7;5:e 06993.
66. De Vries L., Baselmans B., Bartels M. Smartphone-Based Ecological Momentary Assessment of Well-Being: A Systematic Review and Recommendations for Future Studies. *Journal of Happiness Studies*. 2021;22:2361-2408.
67. Ethics and Governance of Artificial Intelligence for Health: WHO Guidance. Geneva, World Health Organization, 2021. Licence: CC BY-NC-SA 3.0 IGO.
68. Ethics Guidelines for Trustworthy Artificial Intelligence. AI Update: EU High-Level Expert Group Publishes Requirements for Trustworthy AI and European Commission Unveils Plans for AI Regulation. 2020. URL: <https://www.jdsupra.com/legalnews/ai-update-eu-high-level-expert-group-73340/>.
69. Finlayson S.G., Bowers J.D., Ito J., et al. Adversarial Attacks on Medical Machine Learning. *Science*. 2019;363;6433:1287-1289. doi: 10.1126/science.aaw 4399
70. Fiorini N., Canese K., Starchenko G., Kireev E., Kim W., Miller V., Osipov M., Kholodov M., Ismagilov R., Mohan S., Ostell J., Lu Z. Best Match: New Relevance Search for PubMed. *PLoS Biol*. 2018;16;8:e2005343.
71. Freeman K., Geppert J., Stinton Ch., Todkill D., et al. Use of Artificial Intelligence for Image Analysis in Breast Cancer Screening Programs: Systematic Review of Test Accuracy. *BMJ*. 2021;374:n1872. doi: 10.1136/bmj.n 1872
72. Freeman K., Dinnes J., Chuchu N., Takwoingi Y., Bayliss S.E., Martin R.N., Jain A., Walter F.M., Williams H.C., Deeks J.J. Algorithm Based Smartphone apps to Assess Risk of Skin Cancer in Adults: Systematic Review of Diagnostic Accuracy Studies. *BMJ*. 2020:368.
73. Gerke S., Minssen T., Cohen G. Ethical and Legal Challenges of Artificial Intelligence-Driven Healthcare. In *Artificial Intelligence in Healthcare*. Academic Press. 2020;295-336.
74. Ghassemi M. Exploring Healthy Models in ML for Health. AI for Healthcare Equity Conference, AI & Health at MIT, 2021. URL: <https://www.youtube.com/watch?v=5uZROGFYfcA>
75. Gillespie N., Lockey S., Curtis C. Trust in Artificial Intelligence. A Five Country Study. Australia, The University of Queensland and KPMG, 2021.
76. Giulietti M., Cecati M., Sabanovic B., Scirè A., Cimadamore A., Santoni M., et al. The Role of Artificial Intelligence in the Diagnosis and Prognosis of Renal Cell Tumors. *Diagnostics*. 2021;11;2:206.
77. Gong D., Wu L., Zhang J., et al. Detection of Colorectal Adenomas with a Real-Time Computer-Aided System (ENDOANGEL): a Randomised Controlled Study. *Lancet Gastroenterol Hepatol*. 2020;5;4:352-361. doi: 10.1016/S2468-1253(19)30413-3
78. Gopal G., Suter-Crazzolara C., Toldo L., Eberhardt W. Digital Transformation in Healthcare – Architectures of Present and Future Information Technologies. *Clin Chem Lab Med*. 2019;57;3:328-335. doi: 10.1515/cclm-2018-0658
79. Guan J. Artificial Intelligence in Healthcare and Medicine: Promises, Ethical Challenges and Governance. *Chin Med Sci J*. 2019;34;2:76-83. doi: 10.24920/003611
80. Habli I, Lawton T, Porter Z. Artificial Intelligence in Health Care: Accountability and Safety. *Bull World Health Organ*. 2020;98;4:251-256. doi: 10.2471/BLT.19.237487
81. Hannun A.Y., Rajpurkar P., Haghpanahi M., et al. Cardiologist Level Arrhythmia Detection and Classification in Ambulatory Electrocardiograms using a Deep Neural Network. *Nat Med*. 2019;25;1:65-69. doi: 10.1038/s41591-018-0268-3
82. Hashimoto D.A., Rosman G., Witkowski E.R., et al. Computer Vision Analysis of Intraoperative Video: Automated Recognition of Operative Steps in Laparoscopic Sleeve Gastrectomy. *Ann Surg*. 2019;270:414e421.
83. Heaven D. Why Deep-Learning AIs are so Easy to Fool. *Nature*. 2019;574;7777:163-166. doi: 10.1038/d41586-019-03013-5



84. *Haifeng J., François C., Qingquan S., Xia H.* AutoKeras: An AutoML Library for Deep Learning. *The Journal of Machine Learning Research.* 2023;6:1-6
85. *He J., Baxter S.L., Xu J., Xu J., Zhou X., Zhang K.* The Practical Implementation of Artificial Intelligence Technologies in Medicine. *Nat. Med.* 2019;25:30-36. doi: 10.1038/s41591-018-0307-0
86. *Hindricks G., Potpara T., Dagres N., et al.* 2020 ESC Guidelines for the Diagnosis and Management of Atrial Fibrillation Developed in Collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The Task Force for the Diagnosis and Management of Atrial Fibrillation of the European Society of Cardiology (ESC) Developed with the Special Contribution of the European Heart Rhythm Association (EHRA) of the ESC. *Eur Heart J.* 2021;42;5:373-498. doi:10.1093/eurheartj/ehaa612.
87. *Hoffman K.M., Trawalter S., Axt J.R., Oliver M.N.* Racial Bias in Pain Assessment and Treatment Recommendations, and False Beliefs about Biological Differences between Blacks and Whites. *Proceedings of the National Academy of Sciences.* 2016;113;16:4296-4301.
88. *Hollon T.C., Pandian B., Adapa A.R., et al.* Near Real-Time Intraoperative Brain Tumor Diagnosis Using Stimulated Raman Histology and Deep Neural Networks. *Nat Med.* 2020;26;1:52-58. doi: 10.1038/s41591-019-0715-9
89. *Islam M.M., Nasrin T., Walther B.A., Wu C.C., Yang H.C., Li Y.C.* Prediction of Sepsis Patients using Machine Learning Approach: A meta-Analysis. *Computer Methods and Programs in Biomedicine.* 2019;170:1-9.
90. ISO/IEC 22989:2022 Information Technology – Artificial Intelligence. Artificial Intelligence Concepts and Terminology.
91. *Jamthikar A.D., Gupta D., Saba L., Khanna N.N., Viskovic K., Mavrogeni S., Laird J.R., Sattar N., Johri A.M., Pareek G., Miner M., Sfrikakis P.P., Protogerou A., Viswanathan V., Sharma A., Kitis G.D., Nicolaidis A., Kolluri R., Suri J.S.* Artificial Intelligence Framework for Predictive Cardiovascular and Stroke Risk Assessment Models: A Narrative Review of Integrated Approaches Using Carotid Ultrasound. *Comput Biol Med.* 2020;126:104043.
92. *Kaur P., Trehan H., Kaur V., et al.* Analysis of Adaptive Neuro-Fuzzy based Expert System for Parkinson's Disease Diagnosis. *Int J Adv Res Ideas Innov Technol.* 2017;3:1120-1127.
93. *Kelly C.J., Karthikesalingam A., Suleyman M., et al.* Key Challenges for Delivering Clinical Impact with Artificial Intelligence. *BMC Med.* 2019;17;1:195. doi: 10.1186/s12916-019-1426-2
94. *Khan B., Hajira F., Qureshi A., Kumar S., Hanan A., Hussain J., Abdullah S.* Drawbacks of Artificial Intelligence and Their Potential Solutions in the Healthcare Sector. *Biomedical Materials & Devices.* 2023;1:731-738. doi: 10.1007/s44174-023-00063-2
95. *Kiani A., Uyumazturk B., Rajpurkar P., et al.* Impact of a Deep Learning Assistant on the Histopathologic Classification of Liver Cancer. *NPJ Digit Med.* 2020;3:23. doi: 10.1038/s41746-020-0232-8
96. *Kim H, Goo JM, Lee KH, et al.* Preoperative CT-based Deep Learning Model for Predicting Disease-Free Survival in Patients with Lung Adenocarcinomas. *Radiology.* 2020;296;1:216-224. doi: 10.1148/radiol.2020192764
97. *Kirubarajan A., Taher A., Khan S., Masood S.* Artificial Intelligence in Emergency Medicine: A Scoping Review. *J Am Coll Emerg Physicians Open.* 2019;1;6:1691-1702.
98. *Koushik C.S.N., Choubey S.B., Choubey A.* Application of Virtual Reality Systems to Psychology and Cognitive Neuroscience Research. *Cognitive Informatics, Computer Modelling, and Cognitive Science.* JSBT-CI: Academic Press. 2020:133-147. doi: 10.1016/C2018-0-05313-X

99. *Larson D.B., Harvey H., Rubin D.L., Irani N., Justin R.T., Langlotz C.P.* Regulatory Frameworks for Development and Evaluation of Artificial Intelligence Based Diagnostic Imaging Algorithms: Summary and Recommendations. *Journal of the American College of Radiology.* 2021;18;3:413-424.
100. *Latonero M.* Governing Artificial Intelligence: Upholding Human Rights & Dignity. *Data & Society.* October. 2018 October 10:37.
101. *Le Douarin Y., Traversino Y., Graciet A., et al.* Telemonitoring and Experimentation in Telemedicine for the Improvement of Healthcare Pathways (ETAPES program). Sustainability beyond 2021: what Type of Organisational Model and Funding Should be used? *Therapie.* 2020;75;1:43-56. doi: 10.1016/j.therap.2019.12.009
102. *Lee C.S., Lee A.Y.* How Artificial Intelligence Can Transform Randomized Controlled Trials. *Transl Vis Sci Technol.* 2020;9;2:9.
103. *Lee E.E., Torous J., De Choudhury M., Depp C.A., Graham S.A., Kim H.C., Paulus M.P., Krystal J.H., Jeste D.V.* Artificial Intelligence for Mental Health Care: Clinical Applications, Barriers, Facilitators, and Artificial Wisdom. *Biol Psychiatry Cogn Neurosci Neuroimaging.* 2021;6;9:856-864.
104. *Lin H., Li R., Liu Z., et al.* Diagnostic Efficacy and Therapeutic Decision-Making Capacity of an Artificial Intelligence Platform for Childhood Cataracts in Eye Clinics: a Multicentre Randomized Controlled Trial. *EClinicalMedicine.* 2019;9:52-59. doi: 10.1016/j.eclinm.2019.03.001
105. *Liu J.* Artificial Intelligence and Data Analytics Applications in Healthcare General. Review and Case Studies. CAIH2020. Proceedings of the 2020 Conference on Artificial Intelligence and Healthcare, October 2020, 49-53. URL: <https://dl.acm.org/doi/10.1145/3433996.3434006>
106. *Liu X., Rivera S.C., Moher D., et al.* Reporting Guidelines for Clinical Trial Reports for Interventions Involving Artificial Intelligence: the CONSORT-AI Extension. *BMJ.* 2020;370:m3164. doi: 10.1136/bmj.m3164
107. *Loftus T.J., Filiberto A.C., Li Y., Balch J., Cook A..C, Tighe P.J., Efron P.A., Upchurch G.R., Rashidi P., Li X., Bihorac A.* Decision Analysis and Reinforcement Learning in Surgical Decision-Making'. *Surgery.* 2020;168;2:253-266.
108. *Lopez-Jimenez F., Attia Z., Arruda-Olson A.M., Carter R., Chareonthaitawee P., Jouni H., et al.* Artificial Intelligence in Cardiology: Present and Future. *Mayo Clin Proc.* 2020;95;5:1015-39.
109. *Lorent M., Maalmi H., Tessier P., Supiot S., Dantan E., Foucher Y.* Meta-Analysis of Predictive Models to Assess the Clinical Validity and Utility for Patient-Centered Medical Decision Making: Application to the Cancer of the Prostate Risk Assessment (CAPRA). *BMC Medical Informatics and Decision Making.* 2019;19;1: Art. 2.
110. *Manne R., Kantheti S.C.* Application of Artificial Intelligence in Healthcare: Chances and Challenges. *Current Journal of Applied Science and Technology.* 2021;40;6:78-89.
111. *Matheny M., Israni T.S., Ahmed M., Whicher D.* Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril. Washington, National Academy of Medicine, 2022. 294 p. ISBN: 978-1-947103-17-7
112. *McGinnis J.M., Williams-Russo P., Knickman J.R.* The Case for More Active Policy Attention to Health Promotion. *Health Affairs.* 2002;21;2:78-93.
113. *Melin P., Miramontes I., Prado-Arechiga G. A.* Hybrid Model Based on Modular Neural Networks and Fuzzy Systems for Classification of Blood Pressure and Hypertension Risk Diagnosis. *Expert Syst Appl.* 2018;107:146-164. doi.org/10.1016/j.eswa.2018.04.023
114. Monitoring the Implementation of Digital Health: an Overview of Selected National and International Methodologies. 2022;70. URL: <https://www.who.int/europe/publications/i/item/WHO-EURO-2022-5985-45750-65816>

115. *Morain S.R., Kass N.E., Faden R.R.* Learning is not Enough: Earning Institutional Trustworthiness through Knowledge Translation. *American Journal of Bioethics*. 2018;18:31-34.
116. *Morley J., Floridi L.* An Ethically Mindful Approach to AI for Health Care. *Lancet*. 2020;395:254-255.
117. *Mulcahy N.* Recent Cyberattack Disrupted Cancer Care Throughout U.S. WebMD. 20 July 2021. URL: <https://www.webmd.com/cancer/news/20210720/recent-cyberattack-disrupted-cancer-care-us>
118. *Nagendran M., Chen Y., Lovejoy C.A., et al.* Artificial Intelligence Versus Clinicians: Systematic Review of Design, Reporting Standards, and Claims of Deep Learning Studies. *BMJ*. 2020; 368:m689. doi: 10.1136/bmj.m689
119. *Nelson K.M., Chang E.T., Zulman D.M., Rubenstein L.V., Kirkland F.D., Fihn S.D.* Using Predictive Analytics to Guide Patient Care and Research in a National Health System. *Journal of General Internal Medicine*. 2019;34;8:1379-1380. doi: 10.1007/s11606-019-04961-4
120. *Nimri R., Battelino T., Laffel L.M., et al.* Insulin Dose Optimization Using an Automated Artificial Intelligence-Based Decision Support System in Youths with Type 1 Diabetes. *Nat Med*. 2020;26;9:1380-1384. doi: 10.1038/s41591-020-1045-7
121. *Ngiam K.Y., Khor I.W.* Big Data and Machine Learning Algorithms for Health-Care Delivery. *Lancet Oncol*. 2019;20;5:262-273. doi: 10.1016/S1470-2045(19)30149-4
122. *Noseworthy P.A., Attia Z.I., Brewer L.P.C., Hayes S.N., Yao X., Kapa S., et al.* Assessing and Mitigating Bias in Medical Artificial Intelligence: the Effects of Race and Ethnicity on a Deep Learning Model for ECG Analysis. *Circ Arrhythmia Electrophysiol*. 2020;13;3:e007988.
123. OECD Framework for the Classification of AI systems, OECD Digital Economy Papers. Paris, OECD Publishing, 2022. 323 p. doi:10.1787/cb6d9eca-en
124. *Okanoue T., Shima T., Mitsumoto Y., Umemura A., Yamaguchi K., Itoh Y., Yoneda M., Nakajima A., Mizukoshi E., Kaneko S., Harada K.* Artificial Intelligence/Neural Network System for the Screening of Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis. *Hepatology Res*. 2021;51;5:554-569.
125. *O'Leary L.* How IBM's Watson Went from the Future of Health Care to Sold Off for Parts. *Slate*, 2022. URL: <https://slate.com/technology/2022/01/ibm-watson-health-failure-artificial-intelligence.html#:~:text=Recently%2C%20Watson%20Health%20was%2C%20essentially,was%20more%20than%20%241%20billion>
126. *O'Neil C.* Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Penguin Random House LLC, New York. 2016. ISBN 9780553418811
127. *Palmer A.* IBM's Watson AI Suggested "Often Inaccurate" and "Unsafe" Treatment Recommendations for Cancer Patients, Internal Documents Show. *DailyMail.com*. 2018. URL: [https://www.dailymail.co.uk/sciencetech/article-6001141/IBMs-Watson-suggested-inaccurate-unsafe-treatment-recommendationscancer-patients.html?ito=email\\_share\\_article-top](https://www.dailymail.co.uk/sciencetech/article-6001141/IBMs-Watson-suggested-inaccurate-unsafe-treatment-recommendationscancer-patients.html?ito=email_share_article-top)
128. *Parikh R.B., Teeple S., Navathe A.S.* Addressing Bias in Artificial Intelligence in Health Care. *JAMA*. 2019;322;24:2377-2378.
129. *Pee L.G., Pan S.L., Cui L.* Artificial Intelligence in Healthcare Robots: a Social Informatics Study of Knowledge Embodiment. *Journal of the Association for Information Science and Technology*. 2019;70:351-369. doi: 10.1002/asi.24145
130. *Phillips M., Marsden H., Jaffe W., et al.* Assessment of Accuracy of an Artificial Intelligence Algorithm to Detect Melanoma in Images of Skin Lesions. *JAMA Netw Open*. 2019;2;10: e1913436. doi: 10.1001/jamanetworkopen.2019.13436



131. *Picard R.W.* Affective Computing. 1997. URL: <https://www.semanticscholar.org/paper/Affective-Computing-Picard/35858c5a69cf765fd2874e96553f693187da6d66#related-papers>
132. *Quaglio G.L., Boone R.* What if we Could Fight Drug Addiction with Digital Technology? EPRS, European Parliament, 2019.
133. *Quer G., Arnaout R., Henne M., Arnaout R.* Machine Learning and the Future of Cardiovascular Care: JACC State-of-the-Art Review. *J Am Coll Cardiol.* 2021;77;3:300-13.
134. *Raichle M.E., MacLeod A.M., Snyder A.Z., Powers W.J., Gusnard D.A., Shulman G.L.* A Default Mode of Brain Function. *Proc Natl Acad Sci US A.* 2001;98;2:676-682.
135. *Rajpurkar P., Irvin J., Ball R.L., et al.* Deep Learning for Chest Radiograph Diagnosis: a Retrospective Comparison of the CheXNeXt Algorithm to Practicing Radiologists. *PLoS Medicine.* 2018;15;11:e1002686.
136. *Reddy S., Allan S., Coghlan S., Cooper P. A.* Governance Model for the Application of AI in Health Care. *J Am Med Inform Assoc.* 2020;27;3:491-497. doi: 10.1093/jamia/ocz192
137. *Reshetnikov A., Abaeva O., Prisyazhnaya N., Romanova T., Sobolev K., Manukyan A.* The Impact of the COVID-19 Pandemic on Burnout Levels among Healthcare Workers: a Comparative Analysis of the Pandemic Period and Post-Pandemic Period. *Heliyon.* 2024;10;17:e36769
138. *Reshetnikov A., Fedorova J., Prisyazhnaya N., et al.* Health Management for Sustainable Development. 2nd World Conference on Smart Trends in Systems. Security and Sustainability (WorldS4); 2018 Oct 30-31; London, UK. New York. Institute of Electrical and Electronics Engineers, 2019;51-56.
139. *Roda S.* Digital Skills for Doctors – Explaining European Doctors’ Position. *J Eur CME.* 2021;10;1:1. doi: 10.1080/21614083.2021.2014097
140. *Sabottke C.F., Spieler B.M.* The Effect of Image Resolution on Deep Learning in Radiography. *Radiol Artif Intell.* 2020;2;1:e190015. doi: 10.1148/ryai.2019190015
141. *Sapci A.H., Sapci H.A.* Innovative Assisted Living Tools, Remote Monitoring Technologies, Artificial Intelligence-Driven Solutions, and Robotic Systems for Aging Societies: Systematic Review. *JMIR Aging.* 2019;2;2:e15429.
142. *Scheetz J., Rothschild P., McGuinness M., Hadoux X., Soyer H.P., Janda M., Condon J.J., Oakden-Rayner L., Palmer L.J., Keel S., van Wijngaarden P. A.* Survey of Clinicians on the Use of Artificial Intelligence in Ophthalmology, Dermatology, Radiology and Radiation Oncology. *Scientific Reports.* 2021;11;1:1-10.
143. *Schulam P., Saria S.* Can you Trust this Prediction? Auditing Pointwise Reliability after Learning. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics.* 2019:1022-1031. URL: <http://proceedings.mlr.press/v89/schulam19a/schulam19a.pdf>
144. *Seyyed-Kalantari L., Liu G., McDermott M., Chen I.Y., Ghassemi M.* CheXclusion: Fairness Gaps in Deep Chest X-ray Classifiers. *BIOCOMPUTING 2021. Proceedings of the Pacific Symposium.* 2021:232-243.
145. *Shah N.D., Steyerberg E.W., Kent D.M.* Big Data and Predictive Analytics: Recalibrating Expectations. *JAMA.* 2018;320;1:27-28.
146. *Shukla S.* Enhancing Healthcare Insights, Exploring Diverse Use-Cases with K-means Clustering. *International Journal of Management, IT & Engineering.* 2023;13;8:60-68.
147. *Schwartz W.B.* Medicine and the Computer: the Promise and Problems of Change. *N Engl J Med.* 1970;283;23:1257-1264.
148. *Schuler A.* Some Methods to Compare the Real-World Performance of Causal Estimators. Ph.D. Dissertation, Stanford University. 2018. URL: <http://purl.stanford.edu/vg743rx0211>



149. *Singer N.* Tech's Ethical "Dark Side": Harvard, Stanford and others want to Address it. The New York Times. 2018. <https://www.nytimes.com/2018/02/12/business/computer-science-ethics-courses.html>
150. *Sit C., Srinivasan R., Amlani A., Muthuswamy K., Azam A., Monzon L., Poon D.S.* Attitudes and Perceptions of UK Medical Students Towards Artificial Intelligence and Radiology: a Multicentre Survey. *Insights into Imaging.* 2020;11;1:14.
151. *Stokes J.M., Yang K., Swanson K., Jin W., Cubillos-Ruiz A., Donghia N.M., MacNair C.R., French S., Carfrae L.A., Bloom-Ackermann Z., Tran V.M., Chiappino-Pepe A., Badran A.H., Andrews I.W., Chory E.J., Church G.M., Brown E.D., Jaakkola T.S., Barzilay R., Collins J.J.* A Deep Learning Approach to Antibiotic Discovery. *Cell.* 2020;180;4:688-702. e13.
152. *Strianese O., Rizzo F., Ciccarelli M., Galasso G., D'Agostino Y., Salvati A., Del Giudice C., Tesorio P., Rusciano M.* Precision and Personalized Medicine: How Genomic Approach Improves the Management of Cardiovascular and Neurodegenerative Disease. *Genes.* 2020;11;7:747.
153. *Sun T.Q., Medaglia R.* Mapping the Challenges of Artificial Intelligence in the Public Sector: Evidence from Public Healthcare. *Government Information Quarterly.* 2019;36;2:368-383. doi: 10.1016/j.giq.2018.09.008
154. *Takhchidi K.P., Gliznitsa P.V., Svetozarskiy S.N., Bursov A.I., Shusterzon K.A.* Labelling of Data on Fundus Color Pictures Used to Train a Deep Learning Model Enhances its Macular Pathology Recognition Capabilities. *Bulletin of Russian State Medical University.* 2021;4:28-33. doi: 10.24075/brsmu.2021.040
155. *Tobore I., Li J., Yuhang L., et al.* Deep Learning Intervention for Health Care Challenges: Some Biomedical Domain Considerations. *JMIR mHealth uHealth.* 2019;7;8:e11966. doi: 10.2196/11966
156. *Triantafyllidis A.K., Tsanas A.* Applications of Machine Learning in Real-Life Digital Health Interventions: Review of the Literature. *J Med Internet Res.* 2019;21;4:e12286. doi: 10.2196/12286
157. *Vashisht R., Jung K., Schuler A., Banda J.M., et al.* Association of Hemoglobin A1c Levels with Use of Sulfonylureas, Dipeptidyl Peptidase 4 Inhibitors, and Thiazolidinediones in Patients with Type 2 Diabetes Treated with Metformin: Analysis from the Observational Health Data Sciences and Informatics initiative. *JAMA Network Open.* 2018;1;4:e181755.

158. *Vellido A.* Societal Issues Concerning the Application of Artificial Intelligence in Medicine. *Kidney Dis.* 2019;5;1:11-17. doi:10.1159/000492428
159. *Voss C., Schwartz J., Daniels J., et al.* Effect of Wearable Digital Intervention for Improving Socialization in Children with Autism Spectrum Disorder: a Randomized Clinical Trial. *JAMA Pediatr.* 2019;173;5:446-454. doi: 10.1001/jamapediatrics.2019.0285
160. *Wang P., Liu X., Berzin T.M., et al.* Effect of a Deep-Learning Computer-Aided Detection System on Adenoma Detection during Colonoscopy (CADE-DB trial): a Double-Blind Randomized Study. *Lancet Gastroenterol Hepatol.* 2020;5;5:343-351. doi: 10.1016/S2468-1253(19)30411-X
161. *Wanless D.* Securing Good Health for the Whole Population. HM Treasury. 2004.
162. *Westergaard D., Moseley P., Sørup F.K.H., Baldi P., Brunak S.* Population-Wide Analysis of Differences in Disease Progression Patterns in Men and Women. *Nature Communications.* 2019;10;1:1-14.
163. *Wicks P., Richards T., Denegri S., Godlee F.* Patients' Roles and Rights in Research. *BMJ.* 2018;362:k3193.
164. *Wiens J., Saria S., Sendak M., et al.* Do no Harm: a Roadmap for Responsible Machine Learning for Health Care. *Nat Med.* 2019;25;9:1337-1340. doi: 10.1038/s41591-019-0548-6
165. *Wijnberge M., Geerts B.F., Hol L., et al.* Effect of a Machine Learning-Derived early Warning System for Intraoperative Hypotension vs. Standard Care on Depth and Duration of Intraoperative Hypotension during Elective Noncardiac Surgery. *JAMA.* 2020;323;11:1052-1060. doi: 10.1001/jama.2020.0592
166. *Wong A., Otles E., Donnelly J.P., Krumm A., et al.* External Validation of a Widely Implemented Sepsis Prediction Model in Hospitalized Patients. *JAMA Intern Med.* 2021;181;8:1065-1070. doi:10.1001/jamainternmed.2021.2626
167. *Yadlowsky S., Hayward R.A., Sussman J.B., McClelland R.L., Min Y.I., Basu S.* Clinical Implications of Revised Pooled Cohort Equations for Estimating Atherosclerotic Cardiovascular Disease Risk. *Annals of Internal Medicine.* 2018;169;1:20-29.
168. *Zech J.R., Badgeley M.A., Liu M., Costa A.B., Titano J.J., Oermann E.K.* Variable Generalization Performance of a Deep Learning Model to Detect Pneumonia in Chest Radiographs: a Cross-Sectional Study. *PLoS Medicine.* 2018;15;11:e1002683.
169. *Zhu H.* Big Data and Artificial Intelligence Modeling for Drug Discover. *Annu Rev Pharmacol Toxicol.* 2020. Jan 6;60:573-589.

Формат 60х90/16, объём 12 усл. печ. л.

Бумага 80 г/м<sup>2</sup> офсетная.

Гарнитура Times New Roman.

Тираж 1000 экз. Заказ № Н489.

Отпечатано в типографии

ФГБУ ГНЦ ФМБЦ им. А.И. Бурназяна ФМБА России.

123098 Москва, ул. Живописная, 46.

Тел.: +7 (499) 190-93-90.

[rcdm@mail.ru](mailto:rcdm@mail.ru), [lochin59@mail.ru](mailto:lochin59@mail.ru)

[www.fmbafmbc.ru](http://www.fmbafmbc.ru)