

«Математикам надо идти в медицину. Что может быть лучше, чем спасти жизни?»

Data Therapy 2, выпуск 1

В первом эпизоде нового сезона подкаста [Data Therapy](#) мы расскажем о применении искусственного интеллекта (ИИ) в медицине. Вместе с нашими гостями — Владимиром Кохом, исполнительным директором по исследованию данных Сбера, и Ольгой Кардымон, научным сотрудником, руководителем группы «Биоинформатика» Института искусственного интеллекта AIRI — ответим на следующие вопросы:

- Как и когда ИИ вообще начали применять в медицине?
- В каких направлениях медицины ИИ сегодня применяется наиболее широко?
- Как ИИ может по фото человека определить наследственные генетические заболевания?
- С какими проблемами сталкиваются специалисты, внедряющие ИИ в медицинскую практику?
- Какое будущее у ИИ в медицине?

Ведущий подкаста — Дмитрий Прусов, директор департамента монетизации данных X5 Group.

- Всем привет, это второй сезон нашего подкаста Data Therapy, и сегодня у нас новый выпуск (00:50) и мы пригласили экспертов. Друзья, представьтесь, пожалуйста, расскажите, кем и где вы работаете, какая ваша профессиональная область?

- Добрый день. Начну я тогда. Меня зовут Ольга Кардымон, я работаю в Институте искусственного интеллекта AIRI. Занимаю позицию научного сотрудника и руководителя группы В-информатика. Соответственно, я развиваю область применения инструментов machine learning касательно геномных данных, белковых данных, и все, что можно решить с большими данными в геномике и В-информатике.

- Да, коллеги, привет. Меня зовут Владимир Кох. Я руковожу группой исследований в Сбербанке в Лаборатории искусственного интеллекта. Моя группа занимается именно исследованием, так скажем, классической медицины. Это все, что связано с анализом и применением методов машинного обучения для обработки медицинских изображений, рентген, КТ и прочее, для обработки медицинских текстов и сигналов.

- Коллеги, спасибо большое за представление. Тема действительно интересная. Давайте, наверное, начну я. Про тему искусственного интеллекта, собственно говоря, тема интересная и уже у всех давно на слуху. Практически во всех сферах промышленности, экономики, бизнеса так или иначе применяется. Но сегодня мы поговорим как раз про ИИ в медицине. И, насколько мне известно, начиная с 2005 года по 20 год по тем исследованиям, которые я смотрел, количество случаев внедрения ИИ именно в медицинские процессы просто катастрофически росло от года к году. Столь значительный рост, наблюдавшийся, собственно говоря, за последнюю пятилетку, если можно так сказать, о чем может говорить? Это ранние технологии развития привнесли больше возможностей и, собственно говоря, появились какие-то конкретные успехи в применении искусственного интеллекта, и открыли ученым возможность для применения и массового внедрения, либо это что-то иное? Можете поделиться своими мыслями?

- Давайте попробую я начать. Я думаю, что в принципе все, что вы перечислили, имеет место быть. Первое, что сама медицина она стала более оцифрована. То есть искусственный интеллект он может работать только там, где есть цифровой след. Соответственно, сейчас повсеместно уже внедрены различные системы автоматизированные, так называемые электронные медицинские карты, которые хранят данные о ходе лечения, о симптоматике пациента, в общем, насколько это возможно, подробно. Соответственно, накопился массив информации, который можно обрабатывать. Второе – это то, что современные методы как раз таки они, и вычислительные, и алгоритмические они позволяют этот массив данных обрабатывать. И, если мы посмотрим, то, конечно, большой бум был именно с точки зрения исследовательских работ. Да, здесь эти данные, объем публикаций от года к году растет просто практически по экспоненте. Но, если смотреть на количество решений, которые получают FDA удостоверения или аналоги европейские, российские, они тоже растут, но, конечно, порядок этих регистрационных удостоверений сильно отличается. Поэтому, я думаю, что именно бум внедрения он будет чуть позже. Сначала наука, а потом практика.

- В принципе, Володя все сказал, тут сложно что-то добавить. Могу только добавить относительно биологических данных, генетических данных. Понятно, что революция в генетике произошла с применением новых мощных машин для секвенирования полного генома. Это не так давно случилось, и за последние 10 лет, то есть где-то с 2010 года по 2022, мы огромными темпами, семимильными шагами накапливали информацию о геномах жителей разных стран, о геномах разных организмов, о геномах всего на свете, что только можно было. И поэтому, накопив эту информацию до нужного объема, чтобы можно было использовать методы машинного обучения, начинают появляться инструменты для ее анализа. С точки зрения В-информатики и геномики это где-то года с 2015, 14, именно machine learning. Классическая В-информатика, конечно же, была и до этого, в какой-то момент, понятно, что уперлись в потолок.

На сегодняшний день сказать, что именно в медицину мы внедряем какие-то решения, я не могу. Потому что есть куча исследований, статей научных, показывающих, что можно использовать искусственный интеллект в разных областях медицинской генетики. Но вот прямо такого широкого спектра инструментов для конкретных клинических задач нет. Есть несколько, которые мы можем обсудить прямо сейчас.

- Насколько я понимаю, это естественное развитие двух составляющих для применения искусственного интеллекта. Во-первых, стало больше именно оцифрованных данных. Это интересно, кстати, оцифруют ли когда-нибудь старые медкарты и рецепты, написанные врачами. Возможно, когда-нибудь мы поймем, что же они пишут, и что они пишут в рецептах и в назначениях. И второе – это развитие технологической базы, которая позволяет это все быстро обрабатывать, и иметь доступ к этой информации.

- Можете поделиться вашим опытом либо, пониманием, точнее, когда вообще в целом встал вопрос о применении искусственного интеллекта в медицине? То есть, грубо говоря, были всевозможные классические методы анализа данных, исследований и все остальное, но в какой-то момент это просто само собой естественным путем, либо была действительно какая-то сверх невозможная задача, к которой был применен технология машинного обучения либо искусственный интеллект и так далее? Можете об этом рассказать? То есть предысторию.

- Здесь история она на самом деле давняя. Если говорить про классические методы экспертной системы, которые кодировали в некие программы поддержки, принятия

расчетных решений, то это началось еще в прошлом веке, когда была первая волна активная применения подобных алгоритмов. И тогда тоже пророчили, что в ближайшее десятилетие, условно говоря, в 70, 60е, что в следующие у нас будет просто врач или какая-то часть медицины заменена этими алгоритмами. Но, понятно, что потом за этим наступила зима так называемая, которая отложила надолго все эти соображения. Сейчас, конечно, исследователи более аккуратны в принципе в своих прогнозах. Поэтому говорить, что здесь что-то, что сделает какую-то революцию, сложно говорить. Но, безусловно, особенно если смотреть какие-то методы, все, что связано с помощниками в анализе медицинских изображений, то здесь это уже используется на мой взгляд достаточно активно и повсеместно. И это как раз таки идея общая она не замещения врача, а она снять с него рутину. То есть рутинная связана с тем, что человек тот же рентгенолог с утра до вечера смотрит на эти картинки, ему надо, взгляд притупляется, его отвлекают и так далее, а ему нужно какие-то простые вещи на картинке, условно простые, безусловно, для профессионала простые, найти, определить на этом изображении. И, конечно, здесь подобные алгоритмы они очень эффективны. Они и раньше существовали, так называемые КАДы, которые, системы немножко другим способом, просто анализом изображения, анализировали что-то, подсвечивали, просто нейросети они делают это лучше, качественно лучше, и местами даже лучше, чем человек. Поэтому эту рутину, безусловно, передают, автоматизируют. И здесь, я бы сказал, что здесь именно больше всего продуктов, дошедших до использования реального. И количество сертификатов на медизделия в этой предметной области больше всего.

- То есть мы говорим в первую очередь о том, что искусственный интеллект – это больше, как некоторый помощник а) в рутинной работе, и помощник в принятии решений обработки огромного количества массива данных. То есть фактически второе мнение, которое упрощает процесс, и оставляет профессионалу больше, наверное, времени на концентрацию и определение конкретного диагноза или еще что-то.

- Конечно. Просто он может решать очень хорошо конкретную задачу. Фактически он, не знаю, искусственный интеллект делает по ряду параметров некоторую суммаризацию для врача, на основе которой он уже делает взвешенное решение относительно пациента. Если говорить про традиционную классическую медицину и лечебно-диагностический процесс.

- У меня сейчас есть несколько такое ощущение, что мы говорим про медицину, и Володя рассказывает про инструменты, что вот там можно в медицине, здесь можно. А я такая: а вот в генетике, а вот в генетике, а вот с генетикой я вам сейчас расскажу, что называется. Так вот, в генетике есть такая классная история, когда по внешнему виду, по внешним признакам можно описать пациента и установить какое-то редкое наследственное заболевание. То есть научно это называется фенотип пациента, описание фенотипа пациента. И врач генетик с этого и начинает прием, с того, что приводят больного ребенка, врач-генетик на него смотрит и говорит, что у него, не знаю, расширенная переносица, поднятые вверх глаза, не знаю, куча других, смещенные уши вниз, еще что-то. Какие-то такие вещи начинает описывать. И на самом деле вот эта описательная генетика она не очень точная. Математики посмотрят на это и скажут: «Ну что это, выше, ниже, шире, уже, что это вообще, о чем вы говорите?» А для врачей это важно, потому что ровно по этим симптомам они и предполагают, какой дальше проверять ген или группу генов, и какое конкретно заболевание искать. И вот компания Face to (11:00) сделала в свое время, лет 5 назад, приложение, которое по изображению пациента давала проценты риска, лет 5 назад, это был 2015 приблизительно, по изображению пациента, по фотографии пациента давала количество, некую степень риска, что у этого пациента то или другое заболевание. Конечно же, работая в лаборатории, мы

это приложение сразу же скачали и друг друга начали фотографировать. Абсолютно здоровые специалисты, дипломированные, явно без проблем с умственной деятельностью, внешне фенотипически вроде обычные люди. И вот у меня в 2015 или 16 году данное приложение показало высокий риск синдрома Дауна. Понятно, что оно ошиблось.

- Боюсь представить, что вы, Ольга, делали с этой информацией в первый момент.

- Ничего. Смеялась. Поняла, что это скорее всего связано с большими глазами, и поэтому приложение было обучено не на достаточном количестве фотографий. Можно было там поиграться, покривляться, получить другие диагнозы. Но, тем не менее, 2022 год, сегодня это приложение реально используют врачи в своей практике, потому что за 7 лет, которые прошли с момента, как только ребята запустили, как идею, тренироваться, пробовать на пациентах, ставили диагнозы, обучали модель, оно дообучилось настолько, что сегодня это приложение показывает, что и я абсолютно здоровая, и коллеги мои абсолютно здоровы внешне. А у врачей они даже сравниваются, соответственно, с этим приложением, как со вторым мнением специалиста. И уже несколько раз на конференциях профессиональных генетического сообщества я слушала доклады о том, кто лучше поставить диагноз, предположительный диагноз, Face to (12:38) или врач-генетик. По-разному, в зависимости от пациента, конечно, и кейса было, но, тем не менее, вот оно – развитие.

- То есть уже сейчас на научных конференциях основная тема – это все-таки versus профессионалы с огромным стажем, или все-таки уже наученная нейросетка, как в этом примере с приложением, которое за 5 лет достаточно хорошо продвинулось. И дерби, кто выиграет в этой схватке.

- Это не то чтобы основная тема. Нет, это не основная тема сейчас, естественно. Все равно основная тема – это пациент, мутация, что делал, схожий пациент. Вот это вот основная тема. Но, тем не менее, все больше и больше докладов, как использовали ту или иную сеть для решения той или иной медицинской задачи вот именно в клинической генетике. Это очень интересно на самом деле. Пусть это не внедрено обширно в практику на сегодняшний день, но это точно уже обсуждается, на это смотрят одним глазом, не доверяют пока совсем. Но, чем больше мы делаем научных исследований и доказательную базу набираем, тем больше шансов, что скоро это будет использоваться в помощь врачу.

- Я добавлю единственное, что вот эта тема человек versus машина, она скорее, как некое представление и шоу. То есть она очень сильно вовлекает, она в моменте создает дискуссию. То есть можно, сколько угодно, слушать скучные доклады с цифрами, но, если ты показываешь и просто, не то что сталкиваешь, но...

- Взбодрить аудиторию.

- Да. Мы на самом деле подобный опыт имели в 19 году на одной радиологической конференции – это очень круто. То есть потому что, особенно для разработчиков, потому что ты видишь, во-первых, удивление врачей, когда это работает, а, во-вторых, максимальное их применение знаний, когда это не работает. Они там так раскладывают, почему это ошиблось. Но это полезно опять же для разработки, безусловно. Помимо какого-то фана, здесь есть очень ценная обратная сторона – это собрать фидбек и понять, что можно улучшить. Потому что, безусловно, жизнь она немножко другая, чем какие-то эксперименты в лаборатории, и это очень ценно.

- Скажите, пожалуйста, мы частично затронули ранее вопрос по поводу широты применения, я слышал, что искусственный интеллект широко используется с точки зрения, например, распознавания фото тех же самых рентгеновских снимков, можете назвать еще несколько ключевых примеров, где это уже прямо максимально продвинуто, используется фактически на ежедневной основе? Может быть, даже, кстати, если есть такая возможность, условно говоря, назвать топ-5 каких-то вещей, где совсем уже хорошо, и где, наоборот, еще требуются, например, усилия для того, чтобы это внедрять, то есть потенциал есть?

- Смотрите, тут еще важно разделить две темы. Первая – это техническая, вторая продуктовая. С точки зрения техники в принципе, по-моему, были уже публикации, что рентген классический – это самое простое изображение, оно двухмерное, одна или две проекции. На этих снимках ИИ выделяет патологии, по-моему, даже лучше, чем врач. Но это опять же с точки зрения техники, технически. Другой вопрос, как это будет реализовано именно в поддержке принятия врачебного решения, в продукте. Здесь важно понимать, что нужно врачу. Потому что иногда исследователи могут показать, да, что-то там хорошо находят, но это на практике не всегда скажется на результативности бизнес-процесса. Поэтому это первое, что важно помнить. А с точки зрения применения, ну да, рентген, также маммография, то есть это исследования, используемые при скрининге рака молочной железы, компьютерная томография. В общем, чем сложнее вид исследования и тип оборудования, тем сложнее в принципе с точки зрения его обработки. И чем больше вариантов этого оборудования. Потому что есть такое, то есть каждое оборудование в принципе сложное оно откалибровано, да, по стандартам, но по-своему. Оно дает разные сигналы. Есть проблема так называемая домен-шифт, когда врач в принципе видит, он может донастроить изображение так, чтобы ему найти нужную какую-то патологию, но для компьютерного зрения, для нейросетки это сложная задача. То есть, если изображения отличаются разного типа оборудования и так далее, то это сложнее. А, если мы добавляем какие-то еще более сложные вещи типа совмещенные виды исследования, позитронно-эмиссионная и компьютерная томография, там уже два сигнала и так далее.

Поэтому потенциал большой. Опять же вернуться, что технологически многое решено уже на текущий момент, но не все реализовано в продукте. То есть у этого еще есть большой потенциал. И многое зависит опять же от того, насколько это реалистично применить в конкретной клинике. Потому что одно дело – собирать данные, другое дело – оцифровать процесс, это тоже важный момент.

- И я тут добавлю, на самом деле очень созвучно с классической имплементацией любого сервиса либо сервиса, построенного на данных, будь то алгоритм либо уже конечный продукт, в какой-то бизнес-процесс компании. То есть можно вообще найти все данные, очистить, собрать, построить удивительно классную модель, но ключевой вопрос – это как раз имплементация. На самом деле ничем не отличается, будь то бизнес, будь то медицина, проблематика ключевая именно эта. Первая имплементация в процессы, где люди годами работали по одним и тем же лекалам, и для себя что-то новое не готовы либо не понимают, как это внедрять, хотя ключевая все-таки, как мы говорили, история в этом, это второе мнение, просто как бы опора, которая снимает с тебя рутину. И как бы страх замены роботами он все-таки остается, как бы смешно это не было.

- Да, я маленький комментарий, мы сейчас говорим про медицину, а есть же еще медицина – это, как некий бизнес или предприятие, которое работает по вполне понятным причинам, в смысле правилам. И здесь как раз методы, например, настройка маршрутизации пациентов, оптимизация этого расписания, то есть, есть куча обеспечивающих организационных процессов – там это работает. То есть там эти системы внедряются, сервисы, и они оптимизируют какие-то вещи. Про это тоже нужно помнить.

- Я, наверное, добавлю чуть-чуть. А вот у меня в генетике, что называется. Можно, конечно, рассказать здесь про известную вещь, как электронные медицинские карты, которые на самом деле очень важны и очень нужны. Вот если сидит, как мы понимаем, есть ли у нас патогенная мутация, то есть мутация, которая вызывает заболевание. На самом деле это проводится достаточно большой объем работы. Сразу люди в лабораториях выделяют ДНК, ДНК секвенируют, то есть прочитывают его последовательность. Дальше садится специалист В-информатик, который это все накладывает на золотой стандарт генома человека здорового, смотрит отличие каждой мутации, и определяет, является ли эта мутация патогенной или она доброкачественная. Потому что не каждая мутация в нашем организме она обязательно приведет к какому-то негативному эффекту. Более того, большинство мутаций, 99,9% мутаций абсолютно доброкачественные в нашем организме. И, чтобы понять, что эта мутация патогенная, на самом деле приходится читать очень много литературы, приходится отслеживать последние публикации, обращаться к большому количеству баз данных. И на сегодняшний день искусственный интеллект, конечно, помогает эту процедуру укоротить. Если раньше, чтобы проверить мутацию, надо было прямо потратить, то есть срок выполнения такого анализа часто в лабораториях до 60 дней стоит. То сейчас уже видны статьи, где пишут, что использовали архитектуру на основе трансформера, модель, которая, по-моему, Биоберд называется, и эта модель Биоберд она способна из статей сразу вытаскивать информацию про нужный ген, отличать его, отличать, знает ли эта статья про нужную мутацию, и, соответственно, уже обращать внимание интерпретатора на то, что, посмотри, что в этой статье написано конкретно про эту мутацию, потому что я для тебя уже все нашел. Тебе не нужно ни гуглить, ни читать текст полностью.

- Простите за попку, прямо (20:57) Железного человека, который делает отсылку в нужный момент, как триггер, срабатывает: «Обрати внимание, пожалуйста, на эти показатели».

- Да, практически. На самом деле это очень удобно, это экономит кучу времени. Конечно же, такие параноики, как я, они все равно сидят и проверяют дальше все равно ручками все остальное. Но, тем не менее, прецедент создан. То есть вот тебе, пожалуйста, дорогой интерпретатор, в помощь инструмент, почитай. Более того, есть уже сервисы, которые коммерциализируют такие решения для генетических лабораторий. Вот это самая простая вещь, которая вроде как уже введена в практику. Назвать ее, что она введена в медицинскую практику и FDA Approved, ну, это не про это немножко. Здесь больше шаманства и науки какого-то, чем вот точно назначений, которые прописаны врачом для использования. Тут каждый на свое усмотрение, по крайней мере, у нас в стране это точно так.

А что хочется сказать, как основной прорыв, и основная крутая вещь, которая произошла, и которая сейчас начинает применяться в практике, но это, наверное, поймут только биотехнологи на сегодняшний день – это, конечно же, инструмент под названием АльфаФолд от компании Deep Mind, который предсказывает третичную структуру белков. Так получилось, что в нашем организме все зависит от взаимодействия белков между собой и белков с какими-то дополнительными молекулами. Соответственно, чтобы сделать лекарство, необходимо подобрать молекулы, которые будут действовать на конкретные белки. Либо найти тот белок, который будет останавливать взаимодействие другого белка. Так как белки они трехмерные, то, соответственно, и взаимодействие может быть абсолютно разным с абсолютно разными белками. А вот узнать эту трехмерную структуру до 2020 года, до конца 20 года было очень сложно, потому что определяется эта структура методом рентген-структурного анализа, у которой есть этап кристаллографии, а получить этот кристалл белка очень сложно. Это дорого, сложно, не

все ручки умеют это делать. Не каждая лаборатория себе в состоянии вообще это позволить, поэтому у нас не очень много трехмерных структур белка на сегодняшний день задепонировано в базы данных. А вот АльфаФолд можно теперь поставить. В июле 21 года вышел открытый код инструмента, мы его тут же в первый же день поставили. Даже, если я не ошибаюсь, это было 15 июля, потому что этот день для нас был просто уау – в открытом доступе АльфаФолд, давайте пробовать. Мы его тут же себе установили, и на сегодняшний день мы можем ввести формулу любого белка, и получить предсказание АльфаФолд. Не все он идеально, конечно, предсказывает, но, тем не менее, он уже решает достаточно много задач, которые связаны с предсказанием структуры белка, с которым ты хочешь поработать. Вот в моей группе, например, АльфаФолд, и не только АльфаФолд, есть еще схожие инструменты, АльфаФолд, (23:43), ОмегаФолд вышел в этом году, эти все инструменты на сегодняшний день мы используем для того, чтобы предсказывать, как изменяется структура белка в зависимости от появления новых мутаций у вируса. Меняется ли она, теряются ли места, куда садятся антитела. Как будет взаимодействовать антитело с этим белком, если появляются новые мутации.

Для чего это все нужно? То есть, кажется, сейчас я вас просто рассказываю какую-то фундаментальную науку, которая что имеет общего с практикой? Но на самом деле, чтобы лечить пациентов, нам нужен препарат. Моноклональная терапия антителами, соответственно, которой лечили президента США в свое время, у нас пытались сделать, по-моему, так и не доделали на сегодняшний день, она эффективна только тогда, когда мы знаем, какое конкретно антитело мы используем в препарате, и куда конкретно он на белок вируса садится, и как он этот вирус будет останавливать. И вот такие инструменты в биотехнологиях они крайне важны и нужны. А если они еще и автоматизированы, для того, чтобы просто вставил последовательность белка и получил ответ на свой вопрос – это идеальная история, сейчас это невозможно, но хотелось бы, чтобы через 5 лет это случилось. Сейчас, конечно, это сидит специалист, который это все анализирует. Но вот создавать такие инструменты, мне кажется, это правильно, это нужно, это помогает биологам сконцентрироваться на биологических задачах, биотехнологам сконцентрироваться на биотехнологических задачах, соответственно, как сварить дальше этот белок и отправить его в вакцину. Соответственно, иммунологам на иммунологических задачах. А инструмент всю эту рутинную работу по предсказанию, где что сделать и какой белок выбрать, он в принципе уже позволяет нам с той или иной степенью вероятности решать.

- У нас, конечно, аудио подкаст, мне кажется, если бы наши слушатели видели нас, они бы увидели мой открытый рот, потому что это просто фантастически интересная история. То есть я сейчас понимаю, насколько это важно. То есть мы часто говорим именно о каких-то прорывах с точки зрения технологий, в первую очередь в экономических каких-то вещах, но в меньшей степени затрагиваем как раз историю фундаментальную, касательно каждого человека, медицины, исследований, генетики. И это просто потрясающе. Я, если честно, (25:53) что, в принципе, возможно, я на это не обращал внимания, но, я надеюсь, что, прослушав наш подкаст, многие заинтересуются этой тематикой, потому что это просто космос. Извините за эмоциональную составляющую, но я очень надеюсь, что люди заинтересуются этой тематикой. И мы сократим этот путь, о котором мы говорили, не в 5 лет, а, так сказать, пятилетку за 3 года за счет привлечения новых кадров.

Мы поговорили с вами о ключевых векторах развития, текущей фазе, о каком-то приблизительном будущем, но, как в любой истории, нужно обратить внимание на какую-то проблематику. Давайте поговорим про трудности и проблемы, с которыми сталкиваются профессионалы в области использования искусственного интеллекта, будь то в медицине, будь то в генетике. Покидайте основной проблематики. И я бы еще хотел с вами обсудить вопрос комьюнити, то есть насколько оно вообще большое в стране сейчас,

и есть ли какое-то отличие, например, от коллег из других стран? Давайте начнем, наверное, про проблематику, я прямо очень хочу послушать. Потому что я уверен, что в этой непростой сложной теме 100%, наверное, затык на каждом шагу.

- Да, давайте попробую я начать. Про проблемы мы не любим говорить, но попробуем. Смотрите, глобально все неплохо, но есть определенные нюансы. Если говорить про русскоязычный домен, особенно то, что касается анализа медицинских текстов различных, то, конечно, у нас есть дефицит открытых данных. Это с точки зрения продвижения науки и как раз вовлечения, это отсылка к следующему пункту комьюнити, вовлечение в комьюнити. То есть у людей, которым это интересно, они должны иметь субстрат для исследований. То есть, есть модели, есть статьи, которые на англоязычных корпусах это все делают. Но, чтобы воспроизвести это для русского языка, какие-нибудь прогностические модели, прогнозирование какого-то состояния по электронной медицинской карте и так далее, для этого нужны открытые данные. В ряде стран, больших стран, в Китае, в США, в Британии, у них есть массивы данных, которые копятся уже десятилетиями, безусловно, это эволюционирует, не то что это идеальные данные какие-то чистые, они просто, как могли, их копят со временем, и улучшают качество хранения, и они предоставляют их для анализа для исследовательских целей. И это, конечно, очень сильно дает такой буст для развития этой технологии, и в принципе для увеличения сообщества. МЛ и переходы их в область медицины. Я говорю именно про корпус текстовых данных. Если говорить, например, про картинку, и, наверное, про генетику, то здесь у них язык международный, а картинки они и в Африке, грубо говоря, картинки, КТ я имею в виду и рентген, так же, как и последовательность белков в гене и так далее. То есть это проблематика.

Что еще? Отдельно, наверное, можно сказать про регуляторику, опять же про получение доступа и так далее, но этим вроде занимаются. То есть, есть отдельные экспертные группы, на это требуется какое-то время, и ключевое – что этим процессом занимаются.

Про комьюнити могу сказать, что, конечно, оно намного меньше, чем то же европейское и уж тем более стран запада.

- А можно сразу, сорри, что прерываю, мне сразу интересно, почему? Чем это обусловлено? Это вряд ли какая-то фундаментальной базы для комьюнити меньше. Какое ваше мнение, почему так сложилось?

- Смотрите, то есть, даже если посмотреть, давайте идти с конца, то есть, если посмотреть на рынок вакансий тех же в Штатах и, например... в Штатах больше, чем во всей Европе, или что-то такое, просто про вакансии. Соответственно, спрос формирует предложение, люди начинают идти, люди начинают развиваться в этой теме. Это первое. Второе, просто количество людей, наверное, может быть, меньше. Мне сложно сказать, с чем это еще может быть связано. То есть причин, наверное, много, но я вижу, что в первую очередь это количество потенциальных рабочих мест. То есть зачем идти и инвестировать в свое образование, и входить в эти комьюнити, если ты потом не сможешь устроиться. Чем больше будет рабочих мест, тем больше будет точек, где будет использовано ИИ и больше запроса на подобного рода специалистов, безусловно, больше будет и людей, которые в этом участвуют. На самом деле все просто.

- Нет, все абсолютно понятно и логично. действительно, человек вкладывает в себя и в свое образование достаточно много и времени, и средств, для понимания того, в какой момент это окупится. То есть мы работаем не только ради идеи, но и обеспечения своих каких-то потребностей.

- Да, здесь, кстати, есть обратная история. Я читал какую-то публикацию, не знаю, это скорректируется или нет, но пару лет назад в тех же Штатах было снижение количества подачи, короче, количества абитуриентов на рентгенологические специальности. Они боялись, что ИИ их вытеснит, и нет смысла.

- Мы опять возвращаемся, что будут вкалывать роботы, а не человек, и, вместо того, чтобы как раз быть центром принятия решений с помощником, который тебе фактически упрощает твою рутину, все равно страх возникает. Может быть, и оправданный. Потому что, вы сами говорили, что с точки зрения как раз исследования снимков это самое простое, что может быть. Картинки они есть везде.

- Да. Но я бы сказал, что скорее просто появятся, может быть, какие-то более наукоемкие специализации, но навряд ли потребность в рентгенологах она пропадет.

- Ок, услышал. Ольга, может быть, вы о каких-то трудностях расскажете?

- А у меня в генетике...

- Или нюансах, да, или «у вас в генетике» про нюансы.

- Да. На самом деле говорить приходится в первую очередь о доступности данных, потому что все-таки геномные данные – это персональные данные. Чтобы обучать какие-то действительно достоверные модели, нужен большой объем данных. Как правило, это госпрограммы в других странах, они на государственном финансировании абсолютно получают, и эти данные охраняются. И получить доступ к этим данным – вообще это надо 7 кругов ада пройти, чтобы тебе разрешили какой-то кусочек из этих данных для твоего уже проапрувленного исследования использовать и дальше изучать. Это, конечно, большая проблема.

Вторая история, которая тоже с данными связана, она заключается в том, что есть куча генетических лабораторий, разбросанных по всему миру. В виду того, что это персональные данные, то каждый обрабатывает их на персональном сервере. И, казалось бы, что, если бы данные с каждого сервера между собой как-то объединить, скинуться информацией, то в принципе и государственная программа не очень нужна будет, давайте анализировать то, что уже люди сделали за свой счет. Мы наберем большой пул от здоровых людей, от больных людей, описанных людей, соответственно, и можно анализировать. Но, к сожалению, на сегодняшний день это, получается, не очень возможно.

Мне очень нравится, есть два британских стартапа, один из них как раз агитирует на тему того, что давайте объединять лаборатории между собой анонимизированными данными, меняться и искать родственников, то есть это тесты происхождения они делают. Что, если разные лаборатории предоставят данные обращающихся к ним людей, то они лучше найдут связи родственников в разных странах и разных уголках, чем будут оперировать только своими данными. Это первая...

- Сорри, что врываюсь сразу. Мы как раз в одном из выпусков этого, ну, 2 сезона обсуждали как раз технологии распределенных вычислений, блокчейн и все остальное, безопасные, в принципе, мне кажется, это то направление, в которое имеет смысл инвестировать как раз, с учетом того, что огромное количество всевозможных юридических и прочих аспектов с точки зрения ограничений, именно применение таких технологий смогло бы объединить различные лаборатории со всего мира.

- Абсолютно согласна. Вот прямо правильный комментарий. И, главное, что есть уже люди, которые в это смотрят, и пытаются предлагать инвесторам проекты на эту тему. И это очень крутые ребята, с которыми я только в этом году познакомилась. Второй проект он как раз собирается использовать (34:18) объединять данные из разных лабораторий. Не только геномные данные, но и разные биохимические данные, да все, что угодно. Все, что в мокрых лабораториях, у нас есть биология wet и dry. Wet lab – это там, где капают, капают, выделяют ДНК, ставят химические реакции. А сухая биология – это мы, В-информатики и дата сайнтисты, которые в эту область пришли, и все, что накопили, получили и анализируют дальше в цифрах.

И вторая история как раз про (34:47) когда хотят объединить лаборатории в сеть между собой, брать результаты, интерпретировать на своих серверах, дообучать, соответственно, модели, чтобы они лучше показывали в одинаковых тестах воспроизводимость, и возвращать, соответственно, анонимизированно это в те лаборатории, где это все хранится. Очень круто, просто очень круто. Вот у нас проблема, которая решается, что называется, методами искусственного интеллекта. Как можем, так и выживаем, но при этом это вот он, современный мир.

Вторая история, на которую хотелось обратить внимание, уже связанная не с данными, а с компетенциями. К сожалению, я сейчас расскажу очень классную историю, если посадить математика, дата сайнтиста на биологические данные и сказать решить вопрос, как правило, это заканчивается какой-нибудь комедийной историей. Мне на этой неделе рассказали, что посадили математиков смотреть на опухолевые клетки, и он тут же заметил, что опухолевые клетки онкологические они живут только при наличии глюкозы в крови. И первое, что он сказал: «Я знаю, как победить рак. Надо убрать глюкозу из крови». Вот, понимаете, что называется.

- И ведь он прав.

- Отчасти он 100% прав. Это на самом деле проблема... это очень прикольный пример, это проблема любого математика и дата сайнтиста, который изолирован, я терминами бизнеса, изолирован от бизнес-процессов. То есть в вакууме с дата-сетом можно найти все, что угодно, и построить все, что угодно. Но как это применять, особенно когда мы говорим про глюкозу, это, конечно...

- Да. Мы, конечно, уберем глюкозу из крови, опухолевые клетки умрут, но то, что человек умрет при этом тоже – это уже мелочи, что называется.

- Это побочка так называемая.

- Но решение абсолютно логично с точки зрения математики, как ни посмотри. Второй пример был у меня в группе, когда дата-сайнтист математик сказал: «Оля, дай дата-сет, я что-нибудь обучу на нем». И у нас есть папочка, в ней лежат разные дата-сеты. А в генетике отличить последовательность ДНК от последовательности белка, они и те, и другие пишутся буквами, можно только по количеству этих букв. В ДНК 4 буквы, а в белке 21. То есть несколько поразнообразнее будет дата-сет, что называется. И вот он в какой-то момент взял дата-сет, что-то обучил, показывает мне результат, а я понимаю, что вообще что-то непонятное, что-то не так. И я спрашиваю: «Слушай, скажи, пожалуйста, а ты на чем обучался, на ДНК или белке?» «Я обучался на буквах». «Я буквы чьи, ДНК или белка?» Он говорит: «А что, есть разница?» Как бы да, как бы есть разница.

Поэтому я 2 года занимаюсь тематикой внедрения инструментов искусственного интеллекта в биологию, сама по образованию я биолог, и не могу тягаться с моими умнейшими коллегами математиками и дата-сайнтистами в их компетенциях, но я знаю точно: что посадить одного математика на биологическую задачу – он не справится, или

дата сайнтиста, он точно не справится. Он получит результат, но он будет абсолютно без глубокого биологического смысла. И в то же время посадите меня обучать – я даже не сильно к этой задаче смогу пробраться эффективно и успешно. Поэтому нужно обучать специалистов, которые несут в себе оба звена, соответственно, в моем случае это и биологический бекграунд, и математический бекграунд. И, радует меня, что с этого года такие программы даже в России начали появляться. Внимательно наблюдаю, какие специалисты будут выходить по завершении магистратуры, например, по индустриальной В-информатике, которая сейчас существует.

Пожалуй, это основная. Понятно, что есть проблемы с мощностями, но это уже как-то, мне кажется, что все с ними сталкиваются, все про них знают, уже остается просто смириться, что сколько есть, столько есть, на чем есть работать, на том и работаем. Конечно, все это дорого, технически затратно. Но такова специфика нашей работы.

- Можно я сразу уточню относительно как раз специфики работы? Я опять, как всегда, про бизнес. Чтобы показать сначала многополярную Северную звезду, обычно в бизнесе делают так называемые (38:57) и показывают быстроту правильности выбранного пути решения для того, чтобы в эту историю инвестировали больше. Потому что мы сталкиваемся с тем, что для развития нужно хорошее обучение, мощные машины, открытые данные и все остальное. В ваших областях это возможно – показывать какие-то быстрые результаты, чтобы заинтересовать инвесторов, государство, фонды?

- Мы стараемся, что называется. Очень сложно. Потому что, чтобы показывать быстрые результаты, надо уметь ставить конкретные задачи. Если мы говорим про какую-то новую абсолютно задачу, и на нее нет данных, чтобы обучить модель, то, конечно же, лаборатории будут очень долго производить эти данные, для того, чтобы мы потом за 3-4 месяца сделали модель и запустили ее уже в прод. Но такая проблема есть. Это будет дорого, бессмысленно и долго.

Есть второй путь, по которому идем мы. Мы берем открытые дата-сеты, мы берем те последовательности, которые уже известны и депонированы. И, исходя из того, что имеем, решаем интересные нам задачи. Вот так тоже можно, но мы не покрываем при этом всю биологию, которую можно покрыть различными моделями.

- Я добавлю, что да, при желании, конечно, можно все это показать, наверное, вопрос опять же... Можно показать какой-то прототип работающий. Но, если мы говорим про бизнес-эффект, то здесь, понятно, что за 3 месяца, условно, за квартал и даже за полугодие очень сложно это показать.

- То есть МВП здесь не применимо, то есть, как термин, МВП, он вообще в принципе здесь далек?

- Ну почему, можно. Можно говорить про это. Понимаете, тут еще тогда должны сойтись несколько составных частей. Для того, чтобы показать какой-то бизнес-эффект, надо еще тогда найти зрелого заказчика. Кстати, к проблематике, для того, чтобы внедрить технологию какую-то новую, она, как правило, чаще всего связана с каким-то реинжинирингом бизнес-процессов. А для этого, чтобы вообще такую операцию сделать в бизнесе, для этого нужны соответствующие люди. И это уже вообще не про искусственный интеллект, а это опять про классический бизнес. Для этого необходимы в системе здравоохранения люди, которые понимают процессы, которые понимают, изменение этого процесса какой эффект даст, и с этим тоже проблема. То есть глобально сама индустрия она потихонечку подходит к этому, вызревает, так скажем. Есть некоторые коммерческие клиники, с которыми я встречался, которые, просто радуешься и

удивляешься, с ними разговариваешь и думаешь: боже мой, это же крутые ребята, они думают очень круто. А бывает не так, скажем так. Бывает немножко по-другому.

Поэтому для того, чтобы показать научно быстрый результат, наверное, в моей предметной области можно. Особенно на картинке. Но это скорее про лабораторное какое-то использование, показать на ретро данных. Если говорить про реальную МВП, то это сложнее. Но опять же, как говорит Ольга, мы стараемся. Без этого никак. Ну и степ-бай-степ, безусловно. Конечно, в голове мы держим какую-то большую великую цель, а на пути к ней нужно показывать обязательно быстрые победы.

- Хотела еще добавить очень хороший пример. В этом году вышла первая инсилака, искусственная структура белка для вакцины от разных штаммов коронавируса, насколько я помню. Могу ошибаться, здесь, возможно, от гриппа, надо проверить. Одно из наших популярных модных вирусов, что называется. Удивительно, говоря про МВП, удивительно, что саму структуру главного компонента, который вызывает иммунный ответ, сформулировали как раз с использованием АльфаФолда, (42:32) вот этих вот инструментов для структурирования 3Д последовательности белка, 3Д структуры белка. Данную вакцину создали с этим элементом, прокололи мышек, мышки показали какой-то иммунный ответ. Но дальше это все же уходит в клинические испытания, в FDA Approved и так далее. И получается как, мои коллеги дата-сайнтисты, В-информатики свою работу сделали за 3-4 месяца, отдали в мокрую лабу, а дальше мокрая лаба отдает лаборатории с клиническими испытаниями, дальше всякие политические, экономические вот эти все переговоры, и закончится это, наверное, через год тем, что ее либо одобряют, либо не одобряют к использованию. Вот МВП есть, в виде структуры дизайн главного компонента для вакцин есть, вот она вся. А вот что дальше, как ее в практику – вот это все очень долго.

- Мы поговорили про саму проблему, ну как, коснулись не проблемы, а нюансов, давайте так это называть, которые сталкиваются на пути, с которыми вы сталкиваетесь. В заключении хотелось бы поговорить, как всегда, поразмышлять на тему некоторого футуризма, про будущее. Мы обращались часто о том, что есть возможность, точнее, в перспективе мы сократим что-то за 5 лет, и это будет доступнее. Может быть, не знаю, это, конечно, неблагодарное дело – предсказывать, но, тем не менее, это тоже имеет место быть, вы, как профессионалы в своих областях, расскажете, куда стремиться с точки зрения развития искусственного интеллекта, какие области, например, не затронуты. Почему это важно, потому что люди, которые нас будут слушать, мне кажется, смогут для себя почерпнуть массу интересного, и, может быть, кого-то это увлечет здесь и сейчас. Понятное дело, что он услышит, что в рентгенологи лучше уже, наверное, не торопиться, это мы уже сделали вывод. Шутка, это все не так. Но, может быть, какие-то нюансы про будущее раскроете и поделитесь с нашими слушателями?

- Прогнозировать – дело неблагодарное. Поэтому буду исходить, наверное, из того представления, если бы я в настоящий момент заканчивал или думал поступить куда-то, на что бы я обратил внимание. Мы про это уже говорили, что, наверное, все-таки будущее – это за такими специалистами на стыке двух каких-то предметных областей. В частности, у нас это математика и медицина. И, безусловно, знание математики, при этом иметь представление о медицине хотя бы какое-то базовое – это существенный плюс для будущего специалиста. Я думаю, как для Ольги, в смысле с точки зрения применения и решения ее задач, так и для области просто классической медицины в плане изучения построения модели на основе просто обычных данных медицинских. Поэтому сложно говорить, что будет в будущем. Опять же про тех же рентгенологов. Я думаю, что просто рентгенологи они будут обрастать другими экспертными знаниями. Наверное, в их программе появятся те же подходы к методам машинного обучения. Потому что

формулировать задачу для математика должен все-таки человек из предметной области. Поэтому, если врач будет иметь, будет понимать принципы работы машинного обучения, то есть не то, как эту архитектуру или как модель создавать, как ее кодить, а просто понимать, как инструмент. Знаете, у тебя есть молоток, у тебя есть еще что-то, он должен понимать, какие возможности она ему предоставляет, это принципиально важно, то он будет ставить правильные задачи. И в этой части я бы врачам советовал инвестировать свое время в какие-то базовые курсы. Мне кажется, что, наверное, программа поменяется или уже поменялась, что появится какой-то базовый курс и на лечебном деле, и по общим медицинским специальностям. Математикам, конечно, надо идти в медицину. Боже мой, что может быть лучше, чем спасать жизни? Это, мне кажется, очевидно. То есть я убежден в том, что у этого есть большой потенциал. Да даже взять ту же онкологию, например. Очевидно, что выявление на ранних стадиях рака, это уже определенные виды рака они излечимы. Это сложно применить к онкологии, но там ремиссии будут на 20 лет и так далее. То есть это уже не смертный приговор. Поэтому, как минимум, ради этой только задачи, а это одна из самых, вторая, наверное, по количеству причин смертей заболевание, и оно такое достаточно драматичное. Конечно, надо идти и бороться с этим. Это очень интересно. Я думаю, что и денег в этой среде будет больше и больше. И опять же это такая медицина, это общая предметная область, которая интересна всем, независимо от страны и так далее. То есть это международная задача.

- Тут я сразу обеими руками за. И особенно с точки зрения как раз раннего определения онкологии, думаю, так или иначе каждый слушатель в своей жизни сталкивался с людьми, у которых была онкология. И как раз именно тот момент, что достаточно трудно выявить на ранних этапах, приводил к печальным последствиям. Поэтому всеми руками, ногами, всем своим существом всех призываю идти в эту историю.

- Да, и это, помимо такого драматизма, который сейчас мы немножко за счет онкологии навяли, это очень интересно и весело. Опять же чувство юмора у медиков оно, знаете, как сказать, оно закаляет. Поэтому на самом деле это все очень интересно. Но приходите к нам, кстати, мы часто ищем специалистов, которые максимально заинтересованы в этом, пожалуйста.

- Мне прямо хочется пойти в тему онкологии дальше, и пофантазировать на тему будущего, опираясь на настоящее. Сегодня уже есть такой тест, который называется жидкостная биопсия. У пациента берут кровь, и по отдельным клеткам, которые в малом количестве, малой концентрации могут найти в этой крови, определяют, есть ли у пациента опухоль на ранней стадии, либо пошли метастазы, когда уже знают, что этот пациент прооперирован, опухоль удалена и так далее. просто технология не то что будущего, а уже настоящего, но, к сожалению, не общеприемлемая сейчас, потому что есть технологические нюансы. Они технологические как с точки зрения оборудования, процесса подготовки образца, то есть все, что wet lab и оборудования для wet lab, но они сложные с точки зрения В-информатики, и здесь искусственный интеллект очень и очень нужен. Модели должны находить именно вот эти клетки определенные, их классифицировать, понимать, что да, это точно опухолевая клетка и так далее. И я уверена, что в скором времени, если мы фантазируем на будущее, диагностика ранняя рака она в том числе может основываться на этой технологии жидкостной биопсии. Когда пациент может сдавать кровь, как вот обычно мы, насморк, кашель, повышенная температура, пошли и сдали биохимию крови. А также можно добавить себе в чек-лист сдавать кровь на маркеры опухоли в виде клеточных представителей, присутствующих в крови. Будущее, да, основано на уже настоящем, то есть уже есть такие тесты, которые под отдельные типы онкологии применяются на западе.

Второе, что мне очень хочется, недавно мы разговаривали с представителем одной из компаний, которые делают генетику для здоровых, то есть это предрасположенности по жизни к разным заболеваниям. И вот мы хорошо поужинали, хорошо пообщались, очень друг друга уважаем и любим. И я им все время говорю про медицину и про больных, а они мне про здоровых. Я говорю: «Ваши здоровые итак здоровые, давайте про больных». Я про больных. Они говорят: «Но здоровых-то больше, и экономика здесь больше, соответственно, и интересы здесь больше, потому что все хотят быть здоровыми и прожить здоровую долгую жизнь». Но по факту на сегодня предсказание таких тестов не очень точное. То есть оно к тому, что будет ли у вас инсульт, инфаркт в 50 лет, это зависит не только от генетики, а это зависит в том числе и от того, какой образ жизни вы ведете, в какой среде проживаете, соответственно, сколько вы нервничаете, сколько чашек кофе пьете и прочее, прочее. К сожалению, генетика на эти вопросы не отвечает, или к счастью, кто как посмотрит.

Но есть такие группы, как группа Владимира Коха, которая сейчас делает модели для медицины, для создания разных гаджетов, которые могут и отклонения в сердечном ритме определить, и повышенный состав сахара в крови, и прочее-прочее. И все эти гаджеты у нас все любят носить, компания Apple тому пример со своими часами и разными настройками, которые развиваются в приложениях. Чем больше таких гаджетов, и больше генетических тестов вместе объединять, тем ближе и ближе мы приближаемся к высокой точности тестов для здоровых, вот этих предсказательных тестов, что, товарищ, ты что-то как-то слишком долго сидишь на стуле, давай-ка ты спортом займешься, а то у тебя сердечная мышца расслабленная слишком, и прочее-прочее, соответственно.

Поэтому я верю, что, если мы все объединим все свои усилия, математический мозг, медицинский мозг, биологический мозг, соответственно, компьютер (51:39) специалистов, то через какое-то время наше здоровое и долгое будущее будет доступно каждому жителю нашей планеты.

- Я не хотел прерывать относительно разговоров о том, что давайте про здоровых либо про больных. Есть одна категория специалистов, которые говорят: нет здоровых – есть недоисследованные. Поэтому, я считаю, что, если подводить технологическую составляющую, мы проговорили для наших слушателей, что точно история с блокчейном и распределенными вычислениями будет востребована в сфере здоровья, медицины, генетики, и там точно есть потенциал. Точно услышал относительно интернета вещей. Это всевозможные носимые устройства, которые так или иначе помогают как раз уже больше собирать и оцифровывать эти данные. И, собственно говоря, индустрии это нужно. Скорее всего за этим тоже последует некоторый технологический прорыв. Поэтому для наших слушателей, кто именно интересуется технологической составляющей, имеет смысл обратить внимание в таком разрезе.

- Ну что ж, я на самом деле с величайшим удовольствием закрываю этот выпуск. Для меня это было просто фантастически. Это, во-первых, очень круто – то, что вы делаете. Это безумно интересно. Я жалею, что я не медик и очень плохо разбираюсь с цифрами с точки зрения того, чтобы собирать какие-то модельки. Но я уверен, что для слушателей были какие-то интересные факты и направления, не знаю, (53:19) либо в целом интереса к этой истории. Поэтому, Владимир, Ольга, огромное вам спасибо. Прямо от души, как говорится. Отличная беседа.

- Спасибо вам.

- Да, спасибо. Будем надеяться, что количество желающих прийти в эту область, неважно, к нам в команду либо просто в комьюнити, что эта передача поможет.

- Да, я как раз хотел об этом и сказать, что благодаря вам на самом деле и, возможно, частичке нашей, если люди послушают, они поймут, что комьюнити есть, оно очень интересное, оно занимается совершенно фантастическими задачами, которые максимально прикладные для человечества, чего уж там скрывать. И, блин, если ваши ряды будут пополняться, это будет очень круто. Спасибо еще раз вам большое.

- Спасибо вам.

- Спасибо большое. Очень приятно было пообщаться.

- Сегодня мы поговорили, как всегда, о данных. И в этот раз мы проговорили про данные в медицине, генетике. Конечно, за такой маленький выпуск нашего подкаста вряд ли мы могли бы осветить все возможные темы использования искусственного интеллекта и данных в таких сложных областях. Но, наверное, ключевая задача была показать, насколько может быть эффективна технология искусственного интеллекта в медицине, и насколько красноречивые примеры решаемых задач, с которыми сталкиваются наши коллеги. Даже с точки зрения банальной экономики здесь есть огромный потенциал. Поэтому, мне кажется, тем, кто все-таки будет слушать нас, присоединяйтесь к комьюнити. Это интересно и это абсолютно благое дело. Что ж, я еще раз хотел поблагодарить наших гостей, что провели с нами увлекательную познавательную беседу. До новых встреч.