

Доклад 17-02-2017.

«Машинное обучение, искусственный интеллект и экономика будущего»

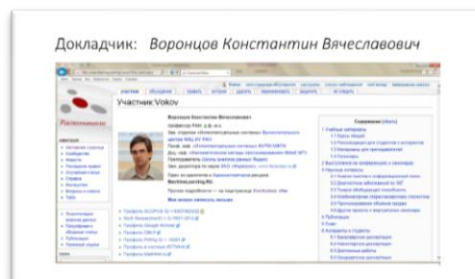
1. Уважаемые коллеги,
2. Я преподаю машинное обучение с 2004 года на Физтехе, на факультете Вычислительной математики и кибернетики МГУ и в Школе анализа данных Яндекс. Много работаю со студентами и аспирантами, под моим руководством защищено более 150 выпускных работ, 10 кандидатских диссертаций. Заведующий отделом «Интеллектуальные системы» Вычислительного центра Российской академии наук. Научные интересы – анализ текстов, информационный поиск, машинное обучение. Полная информация обо мне есть на странице вики-проекта MachineLearning.ru.

3. Мой доклад несёт два основных сообщения.
Первое: технологии машинного обучения, искусственного интеллекта и анализа данных уже меняют нашу жизнь, их роль в экономике будет только расти, они составляют основу цифровой экономики.
Второе: развитые страны, в первую очередь США, считают эти технологии стратегически важными. Обсудим проблемы технологического отставания России и пути его преодоления.

4. Согласно теории технологических укладов Кондратьева, в настоящее время человечество переходит к шестому укладу. Первый уклад начала 19 века связывают с использованием энергии воды и развитием текстильной промышленности. Второй уклад – это энергия пара. Третий уклад – электричество и тяжёлое машиностроение. Четвёртый – нефтехимия, полимеры, автотранспорт. Пятый – электроника, связь, информационные технологии. Сейчас мы переживаем очередную технологическую революцию, которая строится, по словам президента Всемирного экономического форума Клауса Шваба, на «вездесущем и мобильном Интернете, искусственном интеллекте и машинном обучении».

5. Принято считать, что в новом технологическом укладе нас ожидает конвергенция четырёх основных групп технологий: Info, Bio, Nano и Cogno. Из них опережающими темпами развиваются информационные технологии, зародившиеся в недрах пятого уклада и составляющие основу цифровой экономики. Развитие остальных трёх немислимо без компьютерного сбора больших объёмов данных и их анализа средствами машинного обучения.

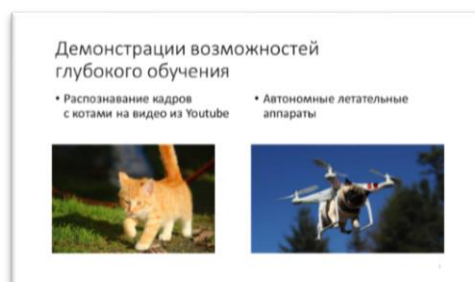
6. Об искусственном интеллекте заговорили в 50-е годы. Первые успехи были связаны с созданием программ для игры в шашки и распознавания трёхмерных сцен. Они были основаны на скрупулёзной формализации знаний экспертов. Постепенно энтузиазм по отношению к экспертным системам остыл, когда выяснилось, что они быстро достигают порога сложности. Когда число экспертных правил достигает нескольких сотен, они начинают конфликтовать и совершать плохо объяснимые ошибки.



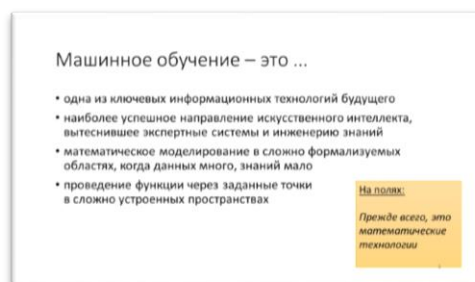
На смену им пришли методы машинного обучения, автоматически настраивающие предсказательные модели по данным. Типичный пример – фильтрация спама в электронной почте.

В последние годы из них выделилось мощное направление – глубокое обучение. Оно позволяет по очень большим данным строить очень сложные модели, причём никакого порога сложности здесь пока не предвидится.

7. Одна из демонстраций возможностей глубокого обучения – конкурсная задача Гугла по автоматическому поиску кадров с котами в потоковом видео на Ютубе, решённая с неожиданно высокой точностью. Понятно, что коты – это лишь рекламный пример. Технологии компьютерного зрения открывают массу новых возможностей: ориентация беспилотных летательных аппаратов в пространстве, распознавание лиц и номерных знаков, автоматизация технологических процессов, предотвращение аварий...



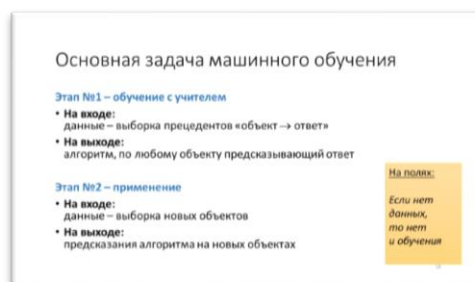
8. Можно по-разному говорить о машинном обучении, и все это правильно. По сути, это восстановление зависимостей по экспериментальным данным. На уроке физики в школе многие делали такой эксперимент: растягивали пружину и измеряли в нескольких точках, как зависит сила растяжения от удлинения пружины. Проводили прямую через полученные точки и проверяли тем самым закон Гука. Машинное обучение – это тот же поиск законов природы по наблюдаемым эмпирическим данным, только зависимости могут быть многомерными и очень сложными, а данные могут быть разнородными, неточными, неполными и даже противоречивыми.



9. Машинное обучение находится на стыке статистического анализа данных и искусственного интеллекта. В науке периодически входят в моду новые направления. Некоторые завершаются мощными прорывами, а другие – осознанием перегретых ожиданий. Неспециалисту трудно бывает разобраться, изменилось ли что-то на самом деле, или новое – это хорошо переименованное старое. Когда я студентом Физтеха пришёл в отдел академика Юрия Ивановича Журавлёва в Вычислительном центре РАН, там занимались распознаванием образов. Потом выяснилось, что аналогичные коллективы за рубежом всё чаще называют свою деятельность «машинным обучением». Потом возник термин Data Mining, трудно переводимый на русский язык. Каждый термин акцентируется на каком-то важном новом аспекте. Например, «большие данные» означает, что на малых данных результат сопоставимого качества недостижим. Большие данные требуют создания технологической инфраструктуры для их накопления и обработки. Однако ни точного определения, ни границ, где кончаются малые данные и начинаются большие, не существует. Термин «наука о данных» мне нравится больше всего. За полвека развития произошла конвергенция нескольких математических и компьютерных дисциплин, и они образовали новую цельную науку.



10. Теперь ещё чуть конкретнее. Одна из основных, и, пожалуй, самая распространённая задача машинного обучения называется обучением с учителем. Она всегда решается в два этапа. На первом этапе нам дают обучающую выборку данных, состоящую из пар «объект, ответ». Предполагается, что существует некий закон природы, зависимость ответа от объекта. Мы должны найти этот закон и описать его конструктивно, то есть построить алгоритм, который по любому объекту выдавал бы как можно более точный ответ.



На втором этапе мы применяем построенный алгоритм, чтобы делать предсказания на новых объектах, которых не было в обучающей выборке.

11. Чтобы показать, что машинное обучение действительно проникло во все сферы жизни, я приведу серию примеров.

1. В задачах медицинской диагностики объектом является запись о текущем состоянии обследуемого, ответом – оценка риска заболевания или рекомендация лечения.

С помощью машинного обучения разрабатываются новые диагностические методики и биомаркеры. Для сбора данных всё чаще используются носимые устройства, это позволяет вести дистанционный мониторинг состояния здоровья.

2. О поиске месторождений золота по геологическим данным я впервые услышал от академика Журавлёва. В 60-е годы по поручению партии и правительства коллектив геологов и математиков решал задачу поиска редких месторождений золота Южно-Африканского типа. Обучающая выборка состояла из 18 объектов – 7 известных месторождений и 11 отрицательных примеров. Объекты описывались 150 признаками. С точки зрения математики тех лет задача считалась нерешаемой и чуть ли не бессмысленной. Тогда Журавлёв изобрёл метод, который применяется в машинном обучении до сих пор. На территории СССР действительно были найдены два месторождения ровно там, где предсказал алгоритм. За эту работу коллектив удостоился Государственной премии. Этот пример показывает, что не только большие данные способны творить чудеса, обучение на малых данных тоже не стоит сбрасывать со счетов, если его грамотно делать.

3. Управление технологическими процессами, казалось бы, очень старая задача, в которой нужно учитывать многолетний опыт специалистов. В 2016 году Фабрика данных Яндекса разработала для Магнитогорского металлургического комбината программу оптимизации расходов ферросплавов и надёжного получения стали с заданным химическим составом. Применяя машинное обучение к данным о выплавках стали, они добиваются экономии ферросплавов на 5%, около 300 млн. рублей в год.

Примеры задач машинного обучения

- **Медицинская диагностика:**
объект – данные о пациенте на текущий момент
ответ – диагноз / лечение / риск исхода
- **Распознавание месторождений полезных ископаемых:**
объект – данные о геологии района
ответ – есть/нет месторождение
- **Управление технологическими процессами:**
объект – данные о сырье и управляющих параметрах
ответ – количество/качество полезного продукта

12. Электронный бизнес и электронная торговля, составляющие основу «цифровой экономики», невозможны без анализа данных и машинного обучения. Вот несколько наиболее известных приложений бизнес-аналитики:

1. Автоматизация кредитования физических лиц началась в США ещё в 70-е годы, до этого кредиты выдавали аналитики по результату личной беседы с заявителем. Оказалось, что компьютеры решают эту задачу не только быстрее, но и точнее людей. Сегодня принятие кредитных решений автоматизировано практически во всех банках.

2. Предсказание ухода клиентов. На высоко конкурентных рынках, например, в сотовой связи, привлечение нового клиента стоит в 50 раз дороже удержания старого клиента. Машинное обучение помогает распознать ранние признаки ухода и понять, к каким абонентам применять маркетинговые воздействия, и какие именно.

3. Прогнозирование объёмов продаж позволяет оптимизировать логистику и сократить издержки на хранение товаров, а в некоторых случаях и вовсе избавиться от складов.

Примеры задач машинного обучения

- **Кредитный скоринг:**
объект – данные о заемщике
ответ – вероятность дефолта, решение о выдаче кредита
- **Предсказание оттока клиентов:**
объект – данные о клиенте на момент времени t
ответ – уйдёт ли клиент к моменту времени $t + \Delta$
- **Прогнозирование объёмов продаж:**
объект – данные о продажах товара на момент времени t
ответ – объём спроса в интервале от t до $t + \Delta$

13. Интернет-компании сегодня являются одними из самых мощных двигателей анализа больших данных.

1. Поисковые компании ежемесячно пополняют свои выборки десятками тысяч экспертных оценок, чтобы улучшать ранжирование результатов поиска.

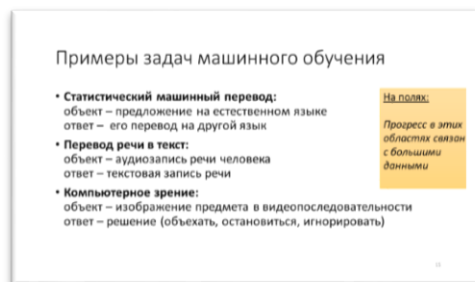
2. Эффективная интернет-реклама вообще невозможна без машинного обучения, поскольку она основана на предсказании вероятности клика пользователя по рекламному баннеру на основании исторических данных.

3. Рекомендательные сервисы делают пользователям персональные предложения о покупке товаров или услуг на основе их прошлого поведения.

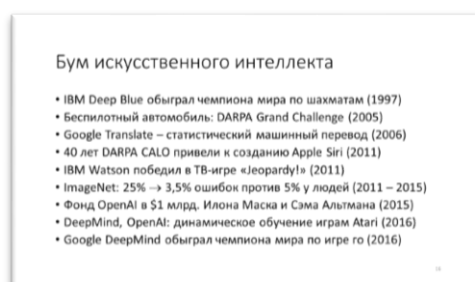
Примеры задач машинного обучения

- **Информационный поиск в Интернете:**
объект – данные о паре «запрос и документ»
ответ – оценка релевантности документа запросу
- **Продажа рекламы в Интернете:**
объект – данные о тройке «пользователь, страница, баннер»
ответ – оценка вероятности клика
- **Рекомендательные системы в Интернете:**
объект – данные о паре «пользователь, товар»
ответ – оценка вероятности, что пользователь купит товар

14. В этих трёх областях машинное обучение и большие данные совершили впечатляющие прорывы в последние годы.
1. Собрав огромную коллекцию параллельных текстов, Гугл смог очень быстро достичь того же качества перевода, которого сообщество компьютерной лингвистики добивалось десятилетиями.
 2. Задача перевода речи в текст ещё лет 10 назад казалась неподъёмной. Делались прогнозы достижения человеческого качества распознавания через 30-50 лет. Сегодня мы имеем голосовой ввод в любом современном смартфоне, и он реально неплох. Правда, лишь на той лексике, по которой разработчики собрали большие данные, но наращивание словаря – это уже простая задача.
 3. Компьютерное зрение – ещё одна область, в которой рост был того же порядка, и прогнозы были весьма пессимистичны, пока лет 5 назад глубокие сети не произвели скачок качества сразу на 10-15%. Для многих задач удалось превзойти человеческое качество распознавания. Без этого прорыва не было бы автономных автомобилей.



15. Современный бум искусственного интеллекта связан с решением ряда задач, которые считались прерогативой человека. Двадцать лет назад компьютер впервые обыграл чемпиона мира по шахматам. В 2005 году агентство DARPA организовало конкурс по созданию алгоритма управления беспилотным автомобилем. Конкурс выиграла команда, которая позже в полном составе перешла в Гугл. В 2011 году компания Apple выпустила приложение для голосового ввода Siri, в основу которого легли результаты 40 лет разработок в рамках проекта DARPA. В том же году система IBM Watson обыграла чемпионов интеллектуальной телевизионной игры, аналога «Своя игра». В том же году стартовал конкурс ИмпиджНет по распознаванию миллиона изображений, и за четыре года удалось добиться точности распознавания лучшей, чем у людей. Основной прорыв произошёл в 2012 году, когда Джеффри Хинтон применил к этой задаче глубокую нейронную сеть, и процент ошибок упал с 25% до 15%. Далее в этом конкурсе лидировали только нейронные сети. Следующее знаковое продвижение – динамическое обучение старым видеоиграм Атари по 4х-часовой видеозаписи. Одна из последних сенсацией – серия выигрышей у чемпионов по игре го, которая всегда считалась сложнее шахмат. Бизнес поверил в то, что эти технологии работают, и началась диффузия инноваций.



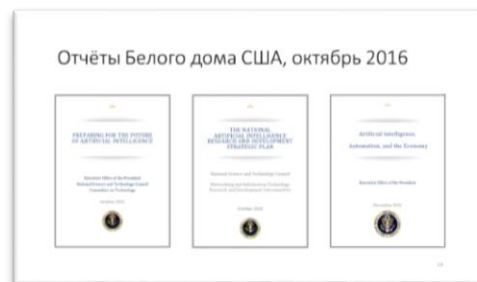
16. Нейронные сети, распознавание сигналов и изображений известны десятки лет назад. Почему же революция произошла лишь в последние годы, и каковы составляющие этого успеха? Во-первых, благодаря тотальной компьютеризации стало возможным собирать и хранить огромные объёмы данных. Во-вторых, одновременно выросли вычислительные мощности, с помощью которых оптимизация моделей по этим данным происходит за приемлемое время. В-третьих, к этому времени накопился значительный опыт в обучении искусственных нейронных сетей и были отобраны наиболее удачные алгоритмы.



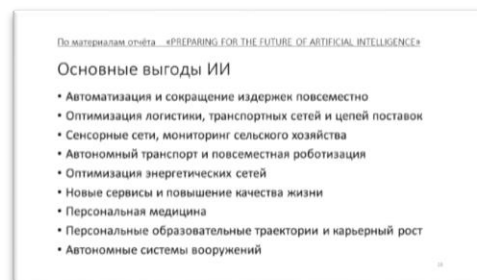
17. В популярной книге Педро Домингоса даётся прекрасный обзор современных применений машинного обучения и основных научных школ. Он считает, что общий искусственный интеллект или «верховный алгоритм», способный решать любые задачи, будет наконец создан, когда удастся объединить все существующие парадигмы в одной технологии. На мой взгляд, это спорное утверждение. Книга ценна прежде всего сотнями примеров того, как технологии машинного обучения проникают во все сферы нашей жизни. И это только начало.



18. В октябре-декабре 2016 года Белый Дом США выпустил ряд отчётов, посвящённых будущему искусственного интеллекта, его влиянию на экономику страны и стратегическому плану исследований и разработок. Достоин внимания процесс подготовки этих отчётов. В мае 2016 года правительство США объявило цели и концепцию широкого обсуждения этих вопросов с научной общественностью. За три месяца Белый Дом организовал пять научных конференций в ведущих научных центрах, в которых приняли участие несколько тысяч учёных. Результаты этих обсуждений были систематизированы в трёх аналитических отчётах, по 30-40 страниц каждый, и выложены в Интернете. Сейчас они доступны свободно.



19. В отчётах отмечается огромная роль искусственного интеллекта в экономике будущего. Компьютерное зрение, голосовой ввод, понимание естественного языка выводят автоматизацию на совершенно новый уровень. Предсказательное моделирование позволяет на всех стадиях бизнес-процессов вводить принципы бережливого производства, а иногда и элементы плановой экономики. Дешёвый сбор и анализ больших данных с помощью сенсорных сетей может в корне изменить управление сельским хозяйством, когда вы получаете возможность на одном мониторе отследить состояние каждой делянки или каждой коровы. Сервисы типа Убера и рекомендательные системы ведут к распределённой экономике, в которой продавец и покупатель связываются друг с другом быстро и напрямую. Ещё одна задача ближайших лет – переход к пожизненному хранению персональных медицинских данных каждого индивида и улучшение клинических практик на основе их анализа. Не менее важная задача – накопление персональной информации об образовании, достижениях, интересах и способностях каждого человека с раннего детства. Технологии социальной инженерии и анализа больших данных можно направить на выявление талантов, персонализацию образовательных траекторий, трудоустройство и формирование коллективов под проекты. Это автоматизированная биржа труда, которая имеет огромный потенциал и способна изменить облик общества.



20. Осознание важности всех этих изменений отражается в постоянном росте государственного финансирования исследований и разработок в области искусственного интеллекта. Однако в отчётах подчёркивается, что 1 миллиард долларов – это слишком мало для США. Сопоставимую сумму инвестирует в технологии искусственного интеллекта маленькая Южная Корея или частные компании, такие как Тойота. Илон Маск инвестировал столько же в небольшую компанию ОпенЭйАй, пригласив в неё пару десятков топовых учёных, в основном в области глубокого обучения.



21. Любая технология является обоюдоострой. Наряду с новыми возможностями, искусственный интеллект несёт в себе потенциальные опасности. Некоторые из них многократно обыграны в голливудских фильмах и произведениях фантастов. Реальная социальная проблема в том, что автоматизация вытесняет низко квалифицированный труд и создаёт новые профессии. Общество может не успевать приспособиться к этим изменениям.



В отчёте разделяются три уровня новых профессий:
— высококвалифицированные учёные и разработчики, их нужно относительно немного;
— инженеры средней квалификации для индустриальных приложений, их нужно очень много;

— потребители и пользователи, занимающиеся обучением систем. Краудсорсинговые платформы Amazon Mechanical Turk и Яндекс.Толока уже сейчас позволяют людям подрабатывать, размечая изображения, тексты или результаты поиска.

22. В отчёте Белого Дома представлены 23 рекомендации.

Нет возможности обсуждать или перечислять их все.

Под номером «один» рекомендация государственным и частным компаниям активнее развивать партнёрство с университетами и научными коллективами для эффективного использования своих данных.

Номер «два» – не бояться открывать свои данные, чтобы в сообществе исследователей не прекращалось соревнование по получению решений максимально возможного качества.

Были и узкоотраслевые рекомендации, например, развивать

систему динамического распределения воздушного трафика. К 2017 году в США зарегистрировано полмиллиона автономных летательных аппаратов, которым уже не хватает воздушных коридоров.

Ещё две довольно очевидные рекомендации – вести мониторинг развития искусственного интеллекта в других странах и активнее финансировать собственные фундаментальные исследования.



23. Отдельный отчёт посвящён национальному стратегическому

плану по исследованиям и разработкам в области искусственного интеллекта. Подчёркивается, что страны с сильной наукой в этой области в будущем займут лидирующие позиции в автоматизации своих экономик.

Предлагается семь основных стратегий:

— долгосрочные инвестиции в научные исследования,

— приоритетное развитие систем, которые не заменяют людей, но работают вместе с людьми,

— выработка этических и юридических принципов работы таких систем в человеческом обществе,

— обеспечение их безопасности, надёжности и высокого уровня доверия людей к ним.

Это всё очень важные, но в целом очевидные стратегии.

Хотелось бы остановиться подробнее на трёх последних стратегиях.



24. Стратегия открытых данных вызывает много сомнений и споров. Кто-то боится конкурентов или «большого брата» или предпочитает исключать риски, упуская возможности. На самом деле выгоды огромны, а риски исключаются при грамотной подготовке данных к раскрытию.

Основные прорывы последних лет достигались именно в конкурсах, и это меняет мышление заказчиков.

Любой отдельно взятый исполнитель всегда заинтересован в том, чтобы зависеть бюджет и занизить требования.

Если ставить задачу как вызов для профессионального

сообщества, то мы быстро получаем оценку предельно достижимого качества решения, находим

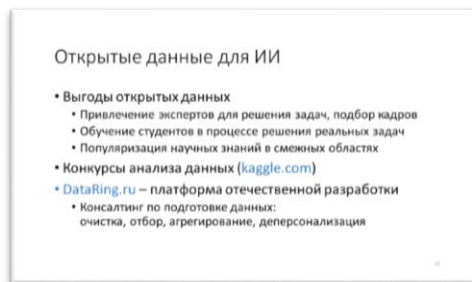
подрядчиков и повышаем рентабельность проекта. Кроме того, конкурсы консолидируют профессиональное сообщество и дают материал для исследований, обучения и популяризации науки.

Один из первых конкурсов в 2006 году организовала американская компания Netflix по прокату видео через интернет. На тот момент 70% ее выручки приходилось на персональные предложения фильмов на сайте. Предложения формировались с помощью машинного обучения, и компания хотела повысить качество своих рекомендаций на 10%. Они назначили приз в миллион долларов — это был первый конкурс со столь крупной суммой. Команды со всего мира соревновались в течение трех лет.

Невозможно нанять такую армию профессионалов всего за миллион долларов. Вслед за этим начался бум конкурсов анализа данных. Сейчас основной конкурсной платформой является Kaggle.com.

Пару недель назад стартовал новый конкурс, который будоражит воображение: предлагается создать искусственный интеллект, который будет отличать фэйковые новости от настоящих.

В России недавно появилась новая платформа DataRing.ru, предоставляющая услуги по подготовке данных и условий конкурсов. Одна из её миссий – формировать культуру проведения конкурсов.



25. Огромные выгоды сулит также использование открытого программного обеспечения. Для разработчиков выгода в том, что они всегда на один шаг впереди остальных. Для сообщества выгод ещё больше: снижение издержек, повышение надёжности кода, выработка отраслевых стандартов, ускорение вывода приложений на рынок. В недавнем конкурсе по алгоритмам автоматического вождения автомобиля выиграли ребята из Яндекса, которые написали всего несколько сотен строк кода поверх свободно доступной библиотеки Гугла TensorFlow.

Открытое программное обеспечение ИИ

- Выгоды открытого кода
 - Снижение издержек, ускорение разработки и внедрения
 - Координация усилий исследователей и разработчиков
 - Снижение технологических барьеров для выхода на рынок
- Примеры открытых экосистем машинного обучения
 - Python + SciPy + Scikit-learn
 - Java + WEKA + RapidMiner
 - Deductor - платформа отечественной разработки (BaseGroup.ru)
- Примеры открытых инструментов для нейронных сетей:
 - DeepMind, OpenAI, TensorFlow, Teano, Torch, Keras, Bonsai, ...

26. SWOT-анализ.

В чём сильные и слабые стороны России в области искусственного интеллекта, на что мы можем повлиять? Каковы возможности и угрозы, на которые мы не влияем? У нас есть сильные научные школы по математике и информатике. Есть компании, признанные лидеры рынка, которые не только развивают технологии и приложения, но и ведут образовательную деятельность. Например, Школа анализа данных Яндекса – это одно из лучших мест в России, где можно за два года овладеть профессией. Высока степень информатизации общества, люди и компании готовы работать через Интернет.

ИИ для России

Сила	Слабость
Возможности	Угрозы

- Научные школы по математике и компьютерным наукам
- Компании-лидеры рынка (Яндекс, АBBYU, Mail.Ru, ...)
- Уровень проникновения Интернета: 97% (16-27 лет), 73% (16 лет и старше)
- Полное проникновение мобильной связи
- Высокая степень компьютеризации

27. Слабых сторон намного больше. Даже сильные стороны – это отчасти слабости, так как всего перечисленного необходимо в десятки или сотни раз больше для перехода к цифровой экономике и новому технологическому укладу.

ИИ для России

Сила	Слабость
Возможности	Угрозы

- Технологическое отставание России
- Доля 5-го уклада в экономике: США 60%, Россия 10%
- Нехватка специалистов, «утечка мозгов»
- Недостаточная технологическая дисциплина
- Недостаточное число дата-центров, их недоступность
- Слабая интегрированность в международное сообщество
- Мало опережающих научных исследований
- Инертная инновационная среда
- Компании неохотно делятся данными

1. В 1988 году в Москве и других крупных городах прошла выставка «Информатика в жизни США». Тогда удивил не уровень компьютерной техники, а степень информатизации американского общества. Уже в восьмидесятых США в целом построили основы пятого технологического уклада, в то время как в СССР отдельные энтузиасты боролись с системой за идеи информатизации. По некоторым оценкам, доля 5-го уклада в экономике России составляет сейчас лишь 10%, тогда как в США 60%.

2. Количество специалистов меньше необходимого не в разы, а на порядки. Увы, самые лучшие уезжают работать за рубеж. Специалист в области искусственного интеллекта приносит своей компании до 10 миллионов долларов в год, в то время как обычный инженер – около 1 миллиона. Немногие компании в России научились столь же эффективно использовать труд этих людей.

3. Внутренняя утечка мозгов не менее страшна, чем внешняя. Люди, которые могли бы быть творцами новых технологий, делают в компаниях достаточно простую работу. Это не только наша беда, таланты остаются недоиспользованными повсюду в мире. Но у нас это сочетается с низкой технологической дисциплиной и культурой коллаборации. Этим навыкам традиционно плохо учат в российских вузах. Отчасти это компенсируется талантами, однако им можно было бы найти и лучшее применение.

4. Двигателем исследований и разработок в области машинного обучения являются сервисы облачных вычислений. В России тоже есть суперкомпьютеры, но облака Амазона доступны просто по Интернету, а для выхода на наши суперкомпьютеры приходится преодолевать бюрократические барьеры.

5. Российское научное сообщество слабо интегрировано в мировую науку по историческим причинам и из-за скудного финансирования. Подготовка хорошей статьи для топовой конференции или научного журнала в нашей области, по оценкам многих моих коллег, стоит около 20 тысяч долларов. Это почти вдвое превышает годовой размер гранта Российского фонда фундаментальных исследований.

6. Мы постоянно догоняем, повторяем опубликованные результаты, значит, отстаём, как минимум, на 2-3 года. Чтобы опережать, необходимы центры компетенций с уникальной творческой атмосферой. Для этого мало выделять финансирование. Необходимо ещё ставить перед ними трудные задачи государственной важности и стимулировать молодёжь оставаться там работать.

7. Большинство аспирантов работают на стороне, так как стипендия – 5 тысяч рублей. Именно в этом причина того, большинство из них не защищают диссертации. В этих условиях очень странно выглядит попытка последних лет обязать аспирантов (взрослых людей) учить ненужные им предметы.

Это отбирает у них последний остаток времени, который можно было бы потратить на исследования.

8. Инертность инновационной среды – это общемировая проблема, связанная с преодолением коммуникационных барьеров между сообществами учёных и предпринимателей. В российских реалиях на это накладывается распределённость и разобщённость профессиональных сообществ.

9. Ещё одна проблема менталитета – непонимание владельцами данных их истинной ценности и выгод от открытых данных и открытого сотрудничества с научно-исследовательскими коллективами. Всё перечисленное не фатально, это слабости, над которыми можно и нужно работать.

28. Возможности – это позитивные факторы, которыми мы не можем управлять.

В первую очередь это стремительный рост мирового рынка анализа данных и искусственного интеллекта.

Доступность научной литературы, онлайн-образования, открытых данных, открытого кода, облачных вычислений, существенно снижают затраты на исследования и разработки. Настолько, что многие стартапы в области интеллектуальных сервисов создаются коллективами из 2-3 человек за 2-3 месяца.

Ещё один фактор, в данном контексте позитивный: труд учёных всё ещё сильно недооценён в России, этим можно и нужно пользоваться.

ИИ для России	Сила	Слабость
	Возможности	Угрозы

- Мировой рынок анализа данных растёт на 40% в год, \$50 млрд. в 2016 году
- Открытые данные и открытый код
- Глобальные облачные сервисы
- Доступность научной литературы
- Доступность платформ онлайн-образования
- Труд учёных в России пока ещё дешёв

29. Угрозы – это негативные факторы, которыми мы не можем управлять.

1. Многие из перечисленных слабостей являются следствием того, что в 90-е мы почти потеряли отечественную науку.

Целое поколение молодых людей не увидели никаких перспектив для себя в российской науке. Сейчас мы имеем демографический провал и разрыв поколений: людей моего возраста в науке удручающе мало. Надеяться приходится на тех, кому сейчас 25-30.

2. Персональные данные российских граждан уже находятся у ведущих американских IT-компаний. Например, всё что вы пишете в гугл-документах, все ваши действия в фейсбуке, медицинские данные с AppleWatch и других носимых устройств могут быть в любой момент проанализированы для выявления вашего статуса и ваших предпочтений. Пока эти данные используются в основном для рекламы товаров, но иные применения также возможны.

3. Британская компания Cambridge Analytica, которая работала на Brexit и предвыборные кампании Теда Круза и Дональда Трампа, использует машинное обучение, чтобы выявлять пол, возраст, расу, социальное положение и психотип человека по его поведению в социальных сетях. Эту информацию они используют для эффективного направленного влияния на мнения людей.

4. Другой пример влияния приводит Джон Перкинс, когда вскрывает механизмы, как американские консалтинговые фирмы навязывают неэффективные дорогостоящие проекты развивающимся странам. Любой консалтинг основан на прогнозах. Если консультант имеет несколько методик прогнозирования, то он выберет ту, которая даст самый оптимистичный прогноз, чтобы зависеть оценку рентабельности проекта и получить заказ. Если заказчик не обладает достаточной экспертизой для перепроверки и всестороннего анализа, то он неизбежно попадает в долговую яму.

5. Зависимость от чужого программного обеспечения также является угрозой для национальной безопасности и независимости страны. Отечественные разработки информационных технологий и искусственного интеллекта должны основываться на открытом коде, который наименее подвержен этой угрозе и позволяет экономить ресурсы, избегая собственных разработок «с нуля».

ИИ для России	Сила	Слабость
	Возможности	Угрозы

- Демографический провал в науке (мало 30-50-летних)
- Глобальные рынки персональных данных уже поделены (Google, Facebook, Amazon, Microsoft, Baidu)
- Прямое влияние на пользователей (Пример: Cambridge Analytica)
- Зависимость от чужих прогнозов (Пример: Джон Перкинс. «Исповедь экономического убийцы»)

30. Технологическое отставание на 30 лет вряд ли возможно, когда технологии обновляются каждые 3 года. Интеллектуализация информационных технологий может стать одним из тех видов деятельности, который позволит нашей стране избежать сценария «ресурсной периферии». В новом технологическом укладе ресурсом становятся знания и таланты. Этот ресурс в стране имеется, но требует более бережного с ним обращения.



Меры в области образования.

1. Поддерживать специализированные учебно-научные центры и университеты, способные готовить отечественную интеллектуальную элиту в области математики и информатики. Бросать их на самофинансирование в угоду концепции Университета 3.0, в условиях незрелой инновационной среды – очень рискованный шаг в российских условиях.
2. Поддерживать создание инженерных курсов по анализу данных. В США есть масса курсов, позволяющих взрослым людям овладевать более современными профессиями. У нас такой практики почти нет. Обучение искусственному интеллекту и анализу данных может быть практико-ориентированным и проводиться на реальных задачах в интересах потенциальных заказчиков.
3. Поддерживать национальную платформу онлайн-образования, которую ведущие университеты уже создают. Это позволит вывести качество онлайн-образования на один уровень с очным образованием.
4. Задействовать средства массовой информации. Сегодня их основная функция – держать население в напряжении и давать негативную картину мира. Чтобы построить цифровую экономику и общество, основанное на знаниях, нужны позитивно мыслящие и информированные граждане. Пора отдать треть эфира просвещению и образованию. Начиная с ликбеза для людей пожилого возраста, как пользоваться банкоматом или сайтом госуслуг, заканчивая пропагандой научных знаний и технологического предпринимательства. Недостаточно создавать возможности. Нужно терпеливо разъяснять населению, как ими пользоваться: где получить современную профессию, как создать стартап, как заработать тысячу рублей, обучая роботов через Интернет, и так далее. Есть масса позитивных примеров успешных российских компаний, в том числе высокотехнологичных. Они должны быть нашими героями, какими для американцев являются Стив Джобс или Илон Маск. У нас таких людей тоже много, и им есть что сказать. Надо чаще предоставлять эфир тем, кто реально движет экономику, науку и инновации.

Меры в области науки и инноваций.

1. Нужны инфраструктурные проекты для преодоления разобщённости профессиональных сообществ. Для начала – общедоступный реестр отечественных разработок с функциями интеллектуального тематического поиска, систематизации знаний, определения трендов и поиска экспертов. Такие проекты затеваются постоянно, но, когда их выполняют отдельные ИТ-компании, общественно значимого продукта почему-то не получается. Чтобы справиться с задачей, необходим консорциум из грантодателей и российских ИТ компаний.
2. Развитием этой идеи является создание всероссийской Интернет-площадки, на которой центры компетенций, заказчики, грантодатели, исполнители и отдельные исследователи смогут быстро находить друг друга, конкурировать и создавать проектные команды с чёткими целями и сроками.
3. Переход к цифровой экономике невозможен без исследований и инноваций в области анализа данных и искусственного интеллекта. Для выработки стратегического плана необходим открытый диалог между властью и учёными. Например, в форме серии научных конференций, как это сделал Национальный совет по науке и технологиям США при подготовке выше упомянутых отчётов.

Меры по формированию рынков данных.

Здесь рекомендации в целом повторяют американский отчёт.

1. Стимулирование партнёрства государственных и частных компаний с научными организациями.
2. Совершенствование законодательной базы по вопросам деперсонализации и раскрытия данных.
3. Развитие облачных сервисов и центров коллективного пользования внутри страны.

При этом надо учитывать, что для развития облачных сервисов сейчас необходима дорогостоящая вычислительная техника и уникальные специалисты, а через 5 это станет совсем дёшево. Будут решены многие инфраструктурные проблемы для обработки потоковых данных. Поэтому на данный момент более рациональная стратегия – сосредоточиться на постановках задач, пилотных проектах, образовательных программах и на самих данных. Можно сделать уже очень много, пользуясь вычислительной инфраструктурой, которую сейчас предоставляют Гугл, Микрософт, Амазон.