

## Лекция 15. Будущее BigData

### Что необходимо сделать

Большие данные уже меняют мир. Нам от них не спрятаться. Они просачиваются в наши страны, города, дома, квартиры и гаджеты. В скором времени эта технология захватит планету.

Тема будущего больших данных очень активно обсуждается по всему миру. Ведь в основе любой коммерческой деятельности лежит информация, как и в основе обеспечения безопасности, и в основе государственного управления. Умные люди обсуждают, как вводить в компьютеры огромные объемы данных и при этом заставить их отыскивать или разрабатывать типовые алгоритмы, недоступные человеку. Нужны суперкомпьютеры для масштабов данных, с которыми человек не справится никогда, с которыми не справляются уже многие современные машины.

Большие данные дают огромные возможности, силу и власть. При этом не следует забывать, что с большой властью и силой приходит или должна придти и большая ответственность. Большие данные — это потенциально большая польза и одновременно большие риски как для пользователей, так и для владельцев систем и баз данных.

Большие данные нейтральны, они ничего не могут сделать сами по себе. Все дело в алгоритмах, а если точнее, то в их качестве. Алгоритмы могут успешно использовать данные или привести какой-то проект к краху. Ведь никому не нужны просто данные, физически много данных, нужно умение правильно соединять эти данные вместе, объединять, разъединять, группировать, принимать решения, делать открытия на основании способности анализировать огромное количество введенной информации. Нужны алгоритмы, способные справляться с гораздо большим количеством информации, чем любой человек. Но нельзя забывать, что алгоритмы не обладают человеческими чувствами, у них нет симпатий и антипатий, а также человеческого здравого смысла. Какую программу заложили — так и работают.

Нужна прозрачность и четкое понимание происходящего, если что-то идет не так. Владельцы систем больших данных должны обеспечить возможность субъектам этих систем видеть, что делают системы, указать на ошибки. Нужно обеспечить возможность внесения исправлений, причем оперативно. Нужны возможности быстрого ввода новых данных и замены старых. Однако многие владельцы больших данных не хотят брать на себя подобную ответственность. Например, утверждают, что алгоритмы являются их собственностью, их работу нельзя объяснять конечным пользователям, а прозрачность принесет вред их бизнесу. Но это неприемлемо! Если система каким-то образом влияет на жизнь людей, нужна защита. А многие системы не приспособлены к корректировке. Это просто не предусмотрено. Корректировка — это не разглашение коммерческих тайн, это обеспечение должной работы системы для правильного использования больших данных во благо, а не во вред. На самом деле системы корректировки не добавляются в алгоритмы из-за стоимости и простой лени.

Например, устанавливается программа, которая находит всех граждан страны, которые зарабатывают свыше одного миллиона долларов в год. Цель — обнаружить тех, кто не платит налоги в больших размерах. Но ведь совершенно необязательно те, кто зарабатывает свыше одного миллиона долларов в год, являются неплательщиками. Подавляющее большинство как раз платит. Это неправильное использование больших данных. А зачем просто так выявлять тех, кто имеет свыше определенной суммы? Это означает неполный, недоработанный алгоритм. Системы больших данных как раз дают

возможность обработать физически большой объем данных с определенными целями. Нужно быть осторожными и максимально правильно интерпретировать результаты.

Если владельцы больших данных будут считать, что эти большие данные — их собственность и никого нельзя к ним подпускать, это примерно то же самое, как не позволять проверять выпускаемые автомобили сторонним лицам. Вы можете себе представить, чтобы владелец автомобилестроительного завода заявил подобное? Как будто бы это раскрытие коммерческой тайны. Вся продукция, выпускаемая в США, проверяется на безопасность. Хотя положение дел с большими данными еще не урегулировано законодательно, как должно бы быть. Это дело будущего. В настоящее время владельцам систем больших данных многое сходит с рук. Например, их алгоритмы решают, что нам продавать, большие данные, в частности, очень помогают маркетологам, которые очень активно используют соответствующие системы для изучения спроса и предложения. Они узнают своих потребителей, привлекают целевую аудиторию из новых потребителей, оценивают их удовлетворенность, реализуют проекты, которые должны пользоваться спросом, усиливают лояльность клиентов, используя новые способы. Но если вы не маркетолог, вы хотите, чтобы они использовали ваши данные? Использовали вас, и вы не знали, как и не могли внести никакую корректировку?

### **Минусы больших данных**

Мы слышим много неприятных вещей о больших данных, хотя, как и в любом деле, здесь есть свои плюсы и свои минусы. Давайте вначале поговорим об отрицательных сторонах больших данных. Самое неприятное возможное последствие, которое я вижу, — это подавление и угнетение. Большие данные в системе государственного управления очень легко могут стать средством контроля за своим населением или просто излишнего контроля и инструментом давления.

Рассмотрим пример Китая. Контроль за населением с помощью системы больших данных вполне может стать реальностью в этой стране. Население в Китае довольно спокойно относится к ограничению личной свободы. Это не США, не Великобритания, не Европа. В настоящее время по приказу китайского руководства ведется разработка системы больших данных, целью которой является автоматическое составление досье на всех граждан, оценка их деятельности, финансов, поведения в целом. Система должна охватить все — траты, кредитоспособность, переход улицы в неполюженном месте, отказ платить в общественном транспорте, нарушение установленных ограничений по количеству детей в семье. Возможно, будут выставляться какие-то оценки или присваиваться баллы. Тем, кто ведет себя правильно, будут доступны награды от каких-то сервисов, контролируемых государством. В зависимости от «благонадежности» граждане будут получать тот или иной уровень услуг при любой деятельности, например заказе блюд в ресторане, знакомствах в Интернете. Если китаец будет действовать правильно с точки зрения государства и демонстрировать благонадежность, все у него будет хорошо. Но он должен правильно вести себя во всем — в общественной и личной жизни! И за этим будет следить система. Но если он что-то сделает не так, то пострадает.

Китайское руководство считает, что такая система необходима, чтобы взять под контроль экономику, где много обмана, где продолжают давать взятки, несмотря на предусмотренные законом суровые наказания, компании то и дело продают продукты питания, которые невозможно есть, или опасные поддельные лекарства. Китайское правительство борется со всем этим и намерено продолжать борьбу. Большие данные помогут решить эти проблемы, но планируемый масштаб предложенной системы охватывает не только мошенников, жуликов и обманщиков.

В Китае в настоящее время живет свыше 1,3 миллиарда человек. Система, охватывающая их всех, все-таки кажется непрактичной, даже при современных технологиях. Она может успешно работать в стране с меньшим населением, но не с таким. Сейчас ее проверяют в региональных пределах. Неизвестно, заработает ли она в масштабах страны, и при этом сложно предсказать все возможные проблемы, которые могут возникнуть при ее применении. Как уже было сказано, при работе с большими данными иногда возникают проблемы, которые нельзя было предусмотреть и предотвратить. Нужно очень тщательно подходить к разработке алгоритмов, ведь подобная система может сломать чью-то жизнь, испортить репутацию. Это представляется страшным для людей проектом.

Да, бывают особые ситуации и особые обстоятельства, когда правительству нужен доступ к данным сверх обычных лимитов, но это должны быть именно особые случаи и особые обстоятельства, а не полный доступ ко всем данным. Нельзя навсегда навешивать на человека ярлык правонарушителя, если он в 18 лет перешел дорогу в неположенном месте, а в 21 поехал на красный свет. У него не должно быть из-за этого проблем в 30 лет при устройстве на работу. А китайская система предусматривает как раз такой подход. Человек не является неблагонадежным в 40 лет, если он чудил в 20. А система больших данных как раз сохраняет эту информацию и может выдать характеристику (плохую) без должных оснований. Станет потенциальный работодатель копать в основаниях, после того как система преподнесла ему вывод? Маловероятно.

Другой неприятный аспект, связанный с большими данными, — хакеры. Мы не хотим, чтобы нашими данными пользовались корпорации, и точно так же не хотим, чтобы ими пользовались хакеры, способные ободрать нас как липку. Они могут разрушить нашу жизнь, взять под контроль «умный» дом, хотя, конечно, их в первую очередь интересуют деньги.

Системы больших данных все больше используются в авиации и больницах. Это иногда пугает. Продвинутые террористы вполне могут проникнуть в базы данных, не взрывая никаких бомб физически, но самолетом станет невозможно управлять, или вдруг исчезнет вся информация о пациентах.

С ростом количества данных должна проводиться работа и по усилению их защиты от террористов и хакеров. Нужно быть бдительными, усиливать безопасность и защиту систем больших данных. Поэтому каждый современный человек должен знать, что такое большие данные и учиться с ними взаимодействовать.

Появилась и новая религия больших данных — датаизм. Ее основные постулаты сформулировал ученый из Израиля Юваль Ной Харари (р. 1976), который учился в Еврейском университете в Иерусалиме, где в настоящее время преподает всемирную историю, и защитил диссертацию в Оксфордском университете. Он прославился книгой «*Sapiens: Краткая история человечества*». Его последняя работа — «*Homo Deus: Краткая история будущего*». В ней он рассказывает о приходе датаизма и об угрозе, стоящей перед человечеством: люди, как мамонты, могут уйти в вечность.

Приверженцы этой «религии» говорят, что за ней будущее, называют ее и идеологией будущего. Основной постулат: Вселенная состоит из потоков данных, и ценность любого явления или субъекта определяется его вкладом в обработку данных. Приверженцы поклоняются большому всемирному потоку данных. Высшее призвание человека, по их мнению, — создать всеобъемлющую систему обработки данных, а потом в нее влиться. Они считают, что при достаточном количестве биометрических данных эта система, обладающая огромной мощностью, способна лучше понять человека, чем он понимает себя сам.

Гуру, как вы догадываетесь, проживают в Силиконовой долине. С одной стороны, это хорошо, с другой — страшно. Расшифровываются тайны человеческого тела, в частности головного мозга, и возникновения чувств. Данные обрабатываются с беспрецедентной мощностью, и в один прекрасный (или ужасный) день появятся системы, которые смогут понимать и контролировать человеческие чувства лучше, чем мы сами. А когда система больших данных станет понимать нас лучше, чем мы сами, власть перейдет от человека к искусственному интеллекту и алгоритмам. Сам человек при этом сольется с непрерывным потоком данных. Хотя нельзя отрицать тот факт, что «датаизм» все больше и больше внедряется в нашу жизнь независимо от того, хотим мы этого или не хотим.

Билл Гейтс, который на протяжении 16 лет возглавлял список самых богатых людей планеты по версии *Forbes*, очень высоко оценил последнюю книгу Харари — об этом он рассказал на своем сайте. Гейтс говорит, что у Харари очень умный взгляд на будущее человечества. Правда, он согласен далеко не со всеми выводами автора, поскольку смотрит на жизнь более оптимистично. Искусственный интеллект становится все более изощренным, но задача людей состоит в том, чтобы он служил человечеству, а не наоборот. Большими данными должны управлять люди.

### **Плюсы**

Если рассматривать большие данные с точки зрения выгоды, то они напоминают «кольцо всевластия» из книги «Властелин колец» Толкиена. Современные технологии позволяют централизованно собирать и обрабатывать данные, обычные люди получают информацию для большей экономической свободы. Мы можем сравнивать цены на интересующий нас товар в режиме онлайн перед тем, как совершить покупку, мы можем делать покупки, не выходя из дома, для этого достаточно коснуться экрана смартфона — и вам все привезут по указанному адресу. Можно работать «на удаленке», то есть онлайн, вести активную социальную жизнь, не посещая никаких мероприятий физически.

Люди массово используют виртуальные платформы для общения и обмена информацией. Все это было невозможно раньше. Вы уже живете в «умном» доме или пока еще нет? Но вы, конечно, слышали, что такие дома уже есть во многих странах, возможно, посещали их и удивлялись «умной» системе освещения — лампочки включаются, когда вы возвращаетесь домой, будто приветствуя вас. И это только одна из функций. Все технологии «умного» дома подключены к базе данных. Это показатель того, как большие данные все больше внедряются в нашу жизнь.

Бизнес получает информацию о предпочтениях пользователей, об их поведении, желаниях и интересах. Данные о пользователях собираются в одно большое информационное поле, обрабатываются и анализируются. Правильная работа с большими данными может принести невероятную выгоду компаниям. Компании уже работают с данными пользователей, примерами могут служить рекламные акции. Но пока никто не знает, что ожидает нас завтра. Какие откроются новые возможности? Какие появятся маркетинговые инструменты?

От больших данных получает пользу и наука. Становится возможным легко получать данные из смежных областей, одновременно анализировать данные, полученные учеными из разных стран и работающих в разных направлениях. Некоторые ученые говорят, что у них сейчас столько данных, что им просто не справиться. Точно так же большие данные помогают и медицине, они помогают отделить факты от догадок, мифы от истинных способов лечения.

В конце концов, мы просто любим получать информацию. Это интересно, это развлекает, приносит удовольствие. Большие данные могут обеспечить невиданный

ранее информационный бум. Правда, необходимым условием является легкий доступ к информации, у нас должны иметься инструменты для управления данными и понимания информации. Мы должны научиться быстро исправлять ошибки и корректировать неправильные алгоритмы.

Следует внести изменения в систему образования — чтобы она учитывала мир больших данных. Нашим детям предстоит жить в мире больших данных, и они должны быть к этому готовы. В настоящее время система образования дает базовые знания для выбранной карьеры, молодым людям во время подготовки к экзаменам приходится заучивать массу информации, которая им потом в жизни не потребуется никогда, а если и потребуется, то ее можно легко найти, а не забивать ею голову. Не нужно запоминать больше необходимого для последующей жизни только для того, чтобы успешно сдать экзамены. Наша система образования строилась в мире, в котором еще не было больших данных, и ориентировалась на жизнь и работу без *Big Data*.

Сейчас возникла необходимость дать молодым людям инструменты для управления данными и понимания их, но при этом сделать так, чтобы система *Big Data* не стала навязчивой идеей, чтобы у молодых людей не появилось новой зависимости. Мы уже знаем про интернет-зависимость, игроманию с зависимостью от онлайн-лотерей с мгновенным выигрышем, мы видим, как молодые люди не могут оторваться от смартфонов и им физически плохо, если они даже на четверть часа лишаются возможности оставаться на связи с многочисленными друзьями. Многие одновременно смотрят телевизор, работают на компьютере и общаются с друзьями через смартфон, иногда в течение одной минуты переключая внимание между тремя источниками. Они не знают, что это опасно для здоровья, в первую очередь психики.

Уроки информатики сейчас уже есть во всех развитых странах, но детей нужно также учить сосредоточивать внимание на одном деле и правильно уметь обращаться с данными. Нужно обучать молодежь навыкам работы с большими данными — как их собирать, обрабатывать и анализировать. Они должны уметь и их корректировать, и распознавать неправильное использование, а в целом — получать максимум пользы от новых технологий. На экзаменах нужно тестировать не память, а навыки. Если не научить молодых людей, которым предстоит жить в мире больших данных, правильному обращению с ними, у человечества в целом могут возникнуть проблемы.

Но давайте все-таки думать позитивно о возможностях, которые открывают перед нами большие данные. Они могут принести нам всем много пользы, если мы научимся их понимать и правильно с ними обращаться, если мы обеспечим прозрачность и возможность корректировки алгоритмов.

На протяжении своей истории человечество сталкивалось со многими открытиями и разработками, которые были и остаются одновременно и позитивными, и негативными, то есть на них можно смотреть как с одной, так и с другой стороны. Например, огонь позволил нашим предкам готовить пищу, обогреваться, но, с другой стороны, пожары уничтожали целые деревни и даже города. То есть если огонь использовать правильно, он приносит пользу, если неправильно, он приводит к трагедиям. Физика дала нам фантастические знания об устройстве Вселенной и технологии, но при этом люди получили возможность быстро и легко уничтожить себе подобных. Энергия атома может быть использована во благо человечества, а может и во вред. Точно так же дело обстоит и с *Big Data*.

Мы не можем сказать: лучше бы этой технологии не было. Она есть. Она существует в том мире, где мы живем. Мы не можем повернуться к ней спиной или жить так, будто ее нет. Большие данные здесь и никуда не исчезнут. Джинн выпущен из бутылки, обратно его загнать не получится, так что нужно думать о том, что пожелать. Нужно научиться желать правильно. У больших данных, как, впрочем, и у

всего в этом мире, есть плюсы и минусы. От нас самих зависит, как мы будем их использовать и насколько светлым будет наше будущее в мире *Big Data*.

**Список использованных источников:**

1. Просто Big Data. — СПб.: Страта, 2019. — 148 с. — (серия «Просто»).