

ОТ НЕСТРУКТУРИРОВАННЫХ ДАННЫХ К ИНСАЙТАМ

О СИЛЕ АНАЛИЗА СОЦИАЛЬНЫХ МЕДИА

Авторы: Линдерт де Воогд, Тара Берд-Ноуланд | Ноябрь 2020

**IPSOS
VIEWS**

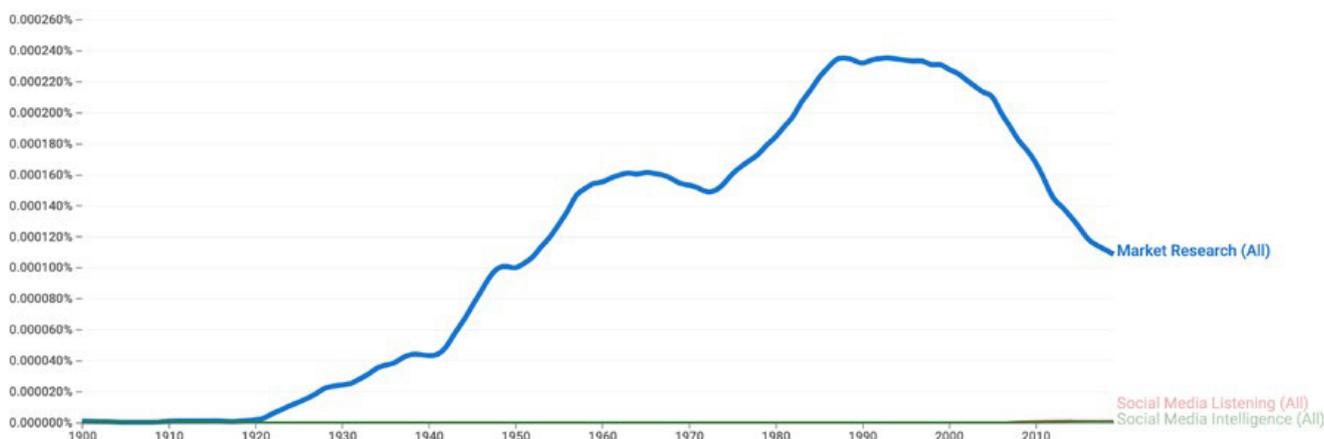
GAME CHANGERS



Вы когда-нибудь представляли себе, что происходило во времена зарождения исследований рынка? Когда исследователи самостоятельно пытались выяснить, что важно, и как наилучшим образом использовать информацию? Согласно Google Ngram Viewer, концепт исследований рынка вошел в широкий лексикон около ста лет назад, и закрепился достаточно быстро (см. Рисунок 1).

Мы стоим в начале пути использования сбора и анализа данных социальных медиа (изначально описываемых как 'social listening', «слушание соцсетей», так как все началось с мониторинга сообщений в социальных медиа). Но фундамент, позволяющий провести на основе этих данных хорошее исследование, виден уже сейчас.

Рисунок 1. Число использований определенных фраз в книгах



Источник: Google Ngram Viewer

Анализ информации из соцсетей (Social Intelligence) – это способность собирать, мониторить и анализировать доступные данные лент соцмедиа (включая социальные сети, блоги, форумы, комментарии и т.д.), чтобы понять, что говорится о какой-либо теме, бренде, организации или явлении.

Концепт Social Intelligence появился на ранней стадии развития социальных медиа. Тем не менее, он берет начало скорее в технологиях, чем в исследованиях. Хотя многие исследовательские агентства (включая Ipsos) обращались к анализу соцсетей еще 10 лет назад, только недавно этот способ работы с информацией стали серьезно рассматривать в качестве настоящей исследовательской дисциплины, а не всего лишь как способ с трудом собрать много деталей, и посмотреть, что люди говорят о брендах. Теперь мы знаем, как социальные медиа и другие неструктурированные наборы данных (объемные и «запутанные» данные с открытыми вопросами, которые не так легко классифицировать) могут рассказать нам полезную информацию последовательно и легко воспроизводимым образом без использования технологии ради технологии.

ОТ ТЕХНОЛОГИЧЕСКОЙ ВОЗМОЖНОСТИ К ИССЛЕДОВАТЕЛЬСКОМУ ПОДХОДУ

За последние 15-20 лет значительно выросшие вычислительные мощности, развитие алгоритмов и искусственного интеллекта (ИИ) позволили нам обрабатывать больше данных, чем раньше. Идет ли речь о платформах для мониторинга соцмедиа, или о многочисленных аналитических подходах, основанных на ИИ, эти достижения в технологиях и анализе данных действительно впечатляют. Они помогают нам использовать и интерпретировать огромные объемы данных быстро и продуктивно. Но им нужно больше, чтобы стать настоящими исследованиями.

Во всей индустрии, организации, от исследовательских компаний до поставщиков технологий, делали маленькие шаги в сторону развития анализа данных из социальных медиа. Ранее компании сферы услуг частично использовали ту силу, которую дают неструктурированные данные, но это было лишь одно из её измерений. Трудность в том, что, как и с количественными и качественными исследованиями, существует великое множество случаев применения анализа данных социальных медиа.

Мы верим, что важно различать то, что является исследованием и то, что им не является. Иногда неструктурированные данные могут быть использованы для неисследовательских целей, в частности, и данные из социальных медиа. Тем не менее,

есть много случаев, когда анализ соцсетей позволял находить ценные инсайты. Поэтому очень важно применять здесь подходы и строгость исследовательских дисциплин, чтобы обеспечить правильность исполнения, учитывая потребности как клиента, так и участников исследования.

Только с применением этих принципов Social Listening будет считаться полноценным исследовательским анализом, обладающим понятной ценностью для клиентов.

Основываясь на многолетнем опыте Ipsos, мы выделили три основополагающих методологических элемента любого значимого исследования социальных медиа: платформы для сбора и анализа данных соцсетей, основанная на ИИ углубленная аналитика и направляемый человеком поиск инсайтов. Сам по себе каждый элемент в отдельности имеет ценность для ответа на важные вопросы клиентов. Но, соединенные вместе, они позволяют получить глубокие инсайты. Как и в любой исследовательской методологии, самой возможности провести исследование недостаточно, для того чтобы сделать это хорошо. Вам нужно иметь необходимые навыки.



Значительно выросшие вычислительные мощности, развитие алгоритмов и ИИ позволили нам обрабатывать больше данных, чем было возможно раньше.

РОЛЬ КАЖДОГО ЭЛЕМЕНТА

Платформы типа Software as a Service (SaaS), будучи самым первым этапом анализа данных соцмедиа, имели больше времени чтобы сблизиться и разойтись в своих подходах. С точки зрения исследований такие платформы – это нечто большее, чем просто инструмент сбора максимального объема данных. Они должны и собирать данные высокого качества, и позволять вам просматривать их продуктивно и эффективно.

Большое разнообразие доступных данных сделало аналитику серьезной зоной роста в секторе технологий. Это огромный объем информации, и человек не может просмотреть ее всю или хотя бы большую ее часть. Нужно добавить аналитику на основе ИИ. Это, действительно, искусственный интеллект, поскольку такая аналитика включает обработку/понимание естественного языка (NLP/NU, natural language processing/understanding), машинное обучение и другие подходы для поиска паттернов и декодирования языка (или картинок/видео). Применение алгоритмов и ИИ к нашим данным помогает нам видеть паттерны. Это не просто использование ИИ ради ИИ, а уверенность, что он направлен на правильные задачи и задействует лучшие возможности, связанные с реальными запросами клиента.

Роль аналитиков здесь нельзя недооценивать. Именно они выбирают и разрабатывают значимые алгоритмы для выполнения отдельных задач. В сфере исследований, ИИ часто рекламируется как «волшебная палочка» из черного ящика, встроенная в платформу. Но, на самом деле, чтобы провести хорошее исследование, вам нужно соединить разные техники. Например, простые векторные модели отлично работают для тематического моделирования, но часто неэффективны для моделирования тональности текста в определенных категориях.

Технологии будут выполнять в точности то, что вы им скажете. Не больше и не меньше. Таким образом, пока платформы SaaS и аналитика на основе ИИ позволяют нам выявлять интересные факты, информацию и паттерны, люди, занимающиеся поиском инсайтов, берут эти кусочки и соединяют вместе, чтобы получить нужные инсайты. В применении исследовательского мышления к тому, что мы делаем, и в использовании проверенных фреймворков, человеку отведена крайне важная роль в нахождении этого понимания.

ТРИ ОСНОВНЫХ ЭЛЕМЕНТА СБОРА И АНАЛИЗА ДАННЫХ В СОЦИАЛЬНЫХ МЕДИА

Платформы для сбора и анализа

данных соцмедиа: Программное обеспечение, спроектированное для возможности сбора данных соцмедиа из разных источников, предоставляющее доступ в реальном времени к различным метрикам при помощи интерактивных дэшбордов.

Аналитика на основе ИИ:

Текстовая, фото и видео аналитика, разработанная для извлечения смыслов из неструктурированных данных при помощи обработки естественного языка (NLP), машинного обучения, data mining, статистического анализа и т.д.

Поиск инсайтов человеком:

Индивидуальный вклад исследователя в поиск инсайтов на основе данных соцмедиа при помощи аналитических фреймворков.

ИЛЛЮСТРАЦИЯ КАЖДОГО ИЗ ЭЛЕМЕНТОВ В ДЕЙСТВИИ

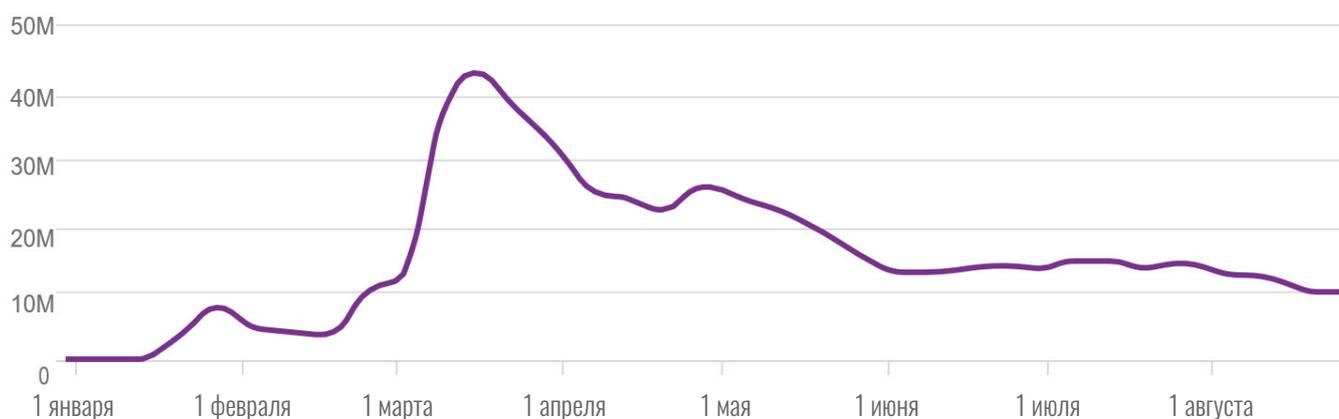
Чтобы проиллюстрировать, что привносит каждый из этих методологических элементов, возьмем для примера тему пандемии COVID-19 и ключевые уроки для брендов. Мы проделали (и продолжаем делать) обширную работу с разными клиентами и типами запросов для лучшего понимания статуса и последствий пандемии.

ПЛАТФОРМЫ ДЛЯ СБОРА И АНАЛИЗА ДАННЫХ СОЦМЕДИА

Используя платформу для сбора данных соцмедиа, такую как Synthesio, мы можем выявить некоторые интересные факты, чтобы понять, что происходит. Для начала, полезно посмотреть, как развиваются обсуждения в сравнении с историческими данными, которые мы также можем собрать уже после наступления событий. Например, к середине мая у нас было больше одного миллиарда твитов только о COVID-19. Спустя три месяца эта цифра удвоилась (см. Рисунок 2).

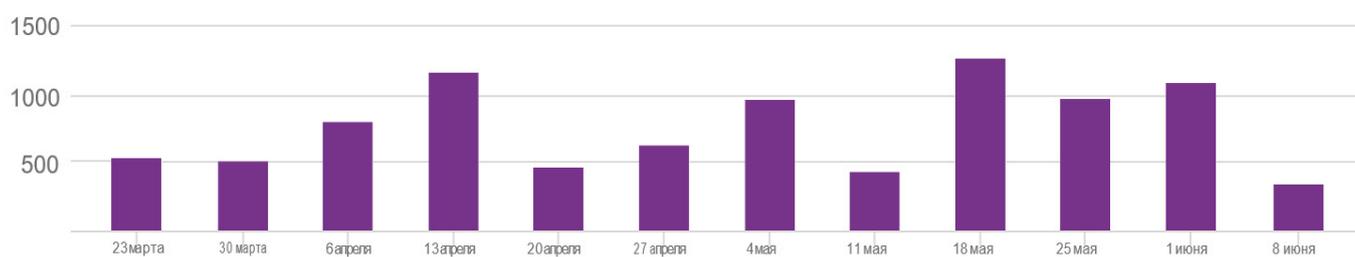
Используя платформу, мы также можем исследовать больше тем, чтобы увидеть на общем уровне, что выделяется, прежде чем начать более глубокий анализ. Такой тип встроенного ИИ, который выделяет аномальные тренды, может сберечь исследователю много времени вместо того, чтобы идти по ложному следу. Например, благодаря функции Сигналы платформы Synthesio, мы можем увидеть, что алкоголь был темой, часто упоминаемой вместе с COVID-19 с апреля по июнь (см. Рисунок 3).

Рисунок 2. Оригинальные посты о COVID-19 в Твиттере в 2020, исключая ретвиты (глобальная база)



Источник: Synthesio, Ipsos

Рисунок 3. Упоминание алкоголя в социальных медиа



Всего упоминаний
9.5K

Всего взаимодействий
99.4K

Число авторов публикаций
6.9K

Потенциальный охват
287.8M

Источник: Synthesio, посты в социальных медиа, написанные на английском языке, март-июнь 2020

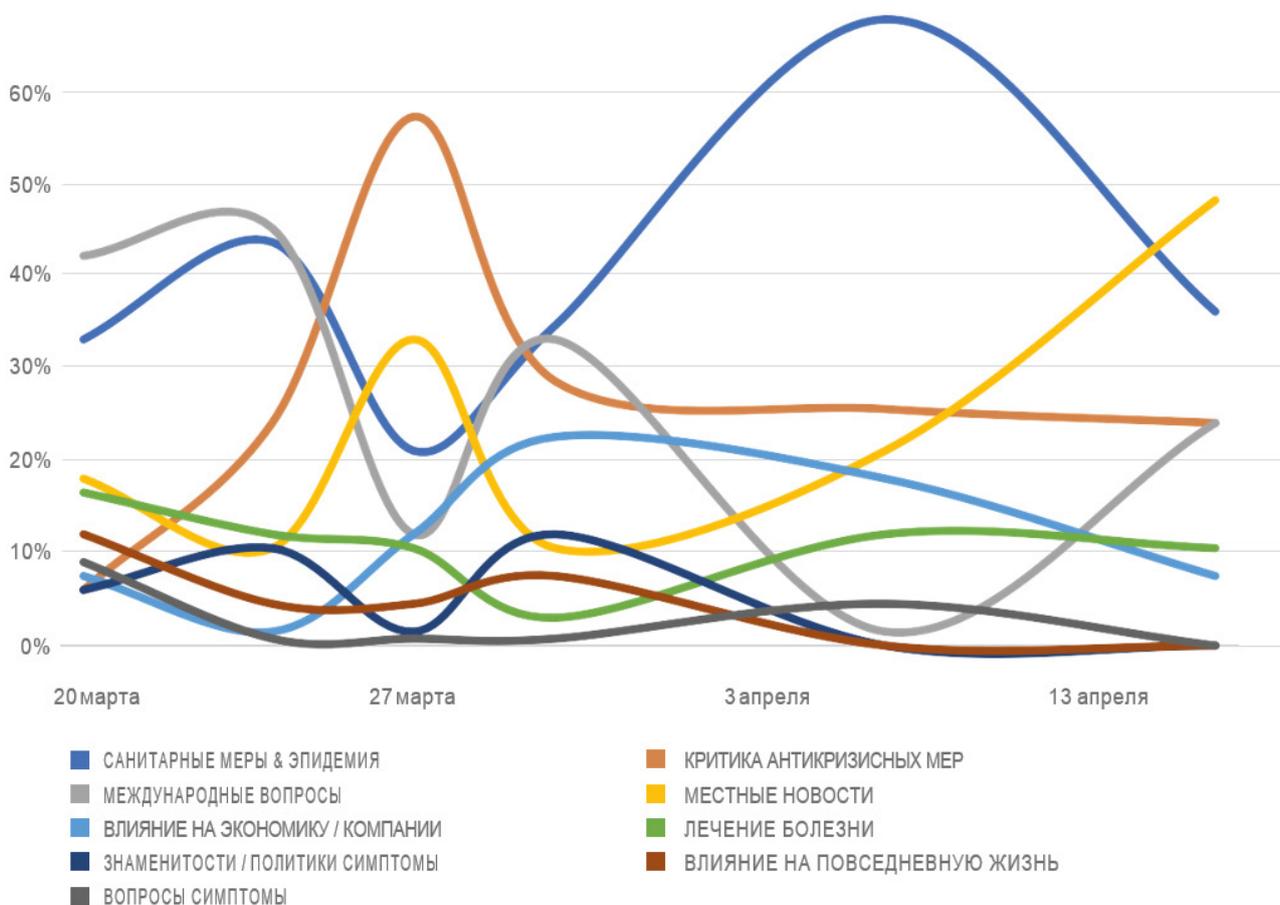


ОСНОВАННАЯ НА ИИ УГЛУБЛЕННАЯ АНАЛИТИКА

В то время как платформы могут предложить первоначальное понимание основных обсуждаемых тем, углубленный анализ на основе ИИ часто необходим для более детализированного понимания истинной природы того, что беспокоит человека, его потребностей и ожиданий. Новые алгоритмы глубокого обучения помогают нам лучше понять, какие темы проявляют себя наиболее ярко, и показывают, есть ли различия восприятия между разными рынками или в пределах одного рынка. Например, на рисунке 4 мы видим, что, во Франции, объемы постов о ключевых темах менялись от недели к неделе (и в некоторых случаях, изо дня в день, как только появлялась новость).

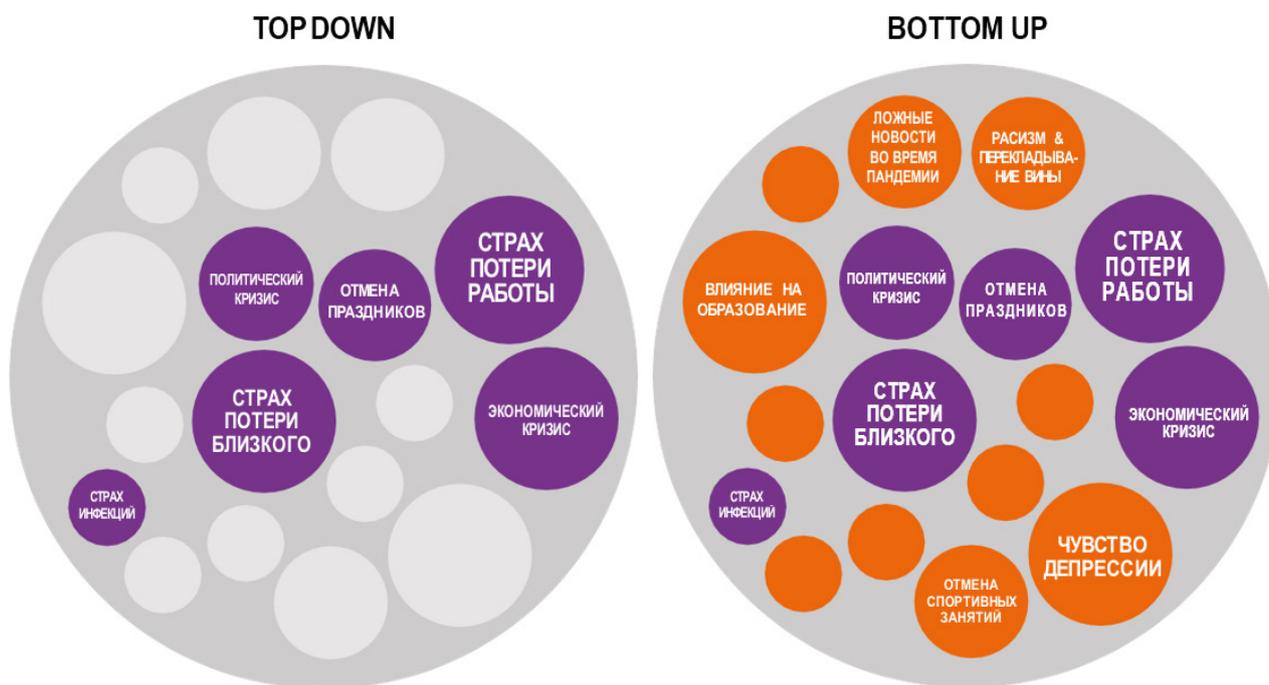
Мы также можем использовать ИИ для понимания эмоций. Здесь мы сфокусировались на страхах людей (см. Рисунок 5), появившихся из-за кризиса COVID-19. Стандартное top-down моделирование состояло бы в определении числа тем в массиве данных. Исследователь может предположить, что люди, на самом деле, были озабочены сохранением работы или страхом потерять близких. Совмещая этот top down подход с подходом, более ориентированным на потребителя и применяющим возможности алгоритмов глубокого обучения (мы называем это bottom-up подходом), мы можем выявить дополнительные новые страхи, которые мы могли не увидеть – такие как негативное воздействие карантина на образование детей и появление расизма или обвинительного поведения во время кризиса.

Рисунок 4. Эволюция тем во Франции (ИИ-генерированное тематическое моделирование)

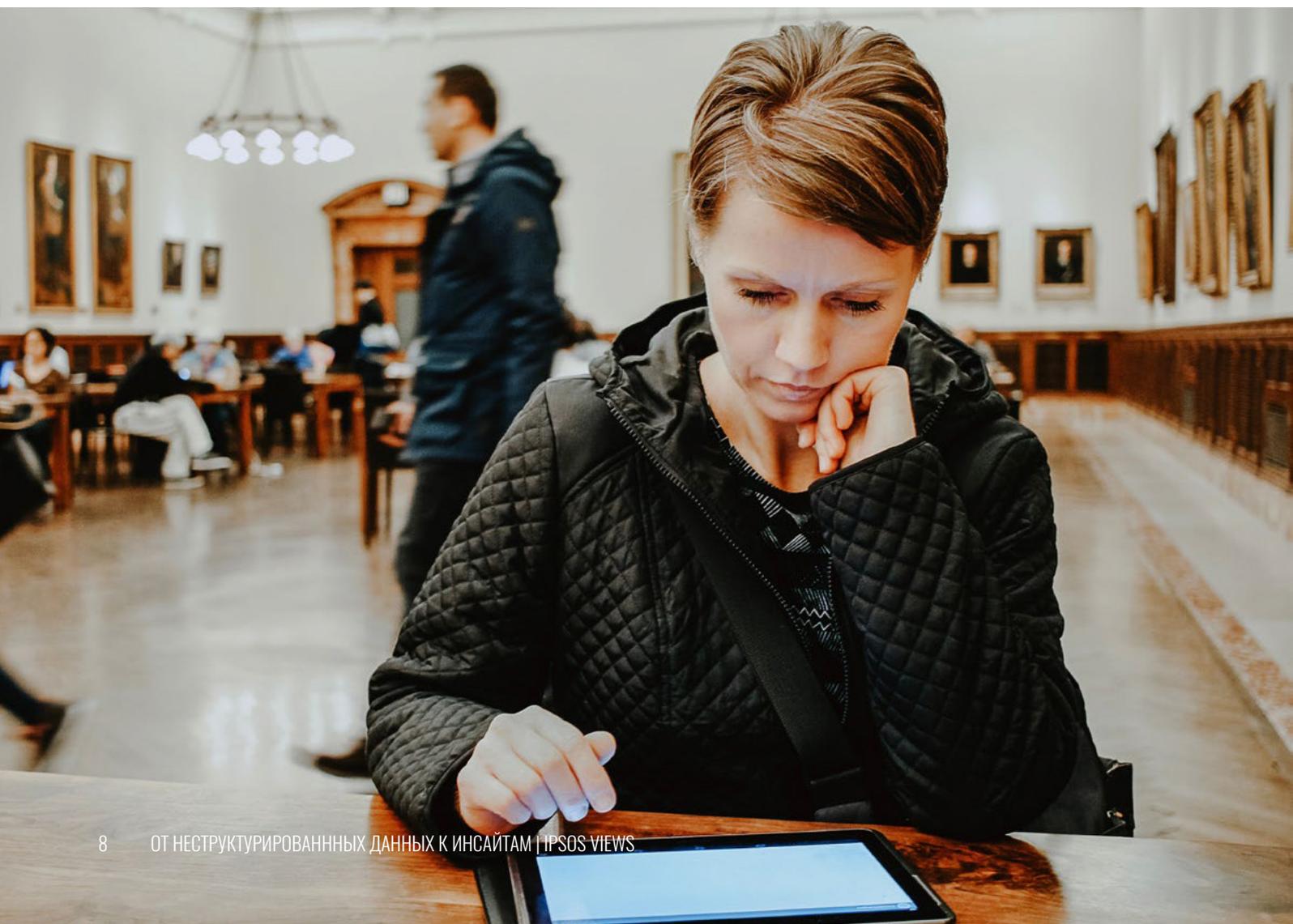


Источник: Сбор и анализ данных социальных медиа, Ipsos, март-апрель 2020

Рисунок 5. Страхи людей: выявление полного спектра



Источник: Анализ данных социальных медиа, Ipsos, посты в социальных медиа, написанные на английском языке



ПОИСК ИНСАЙТОВ ЧЕЛОВЕКОМ

Возможности обнаружения инсайтов человеком позволяют глубже понять ответ на вопрос «почему» происходит то, что происходит, а также оценить последствия изменений потребительского поведения. Это помогает нам понять, что это может значить для брендов

или организаций. Это особенно полезно, когда вы можете связать эти инсайты с другими фреймворками. Например, с фреймворком Ipsos Pandemic Adaptability Continuum 3 (см. Рисунок 6).

Рисунок 6. Ipsos Pandemic Adaptability Continuum



Источник: Ipsos Global Science Organisation

Это помогает нам связывать выводы с целями исследования и объединить их в нечто большее. Например, когда мы изучали изменение привычек питания под влиянием пандемии коронавируса, мы видели, что потребители проходят через разные стадии (см. Рисунок 7). Здесь можно увидеть новые возможности для компании.

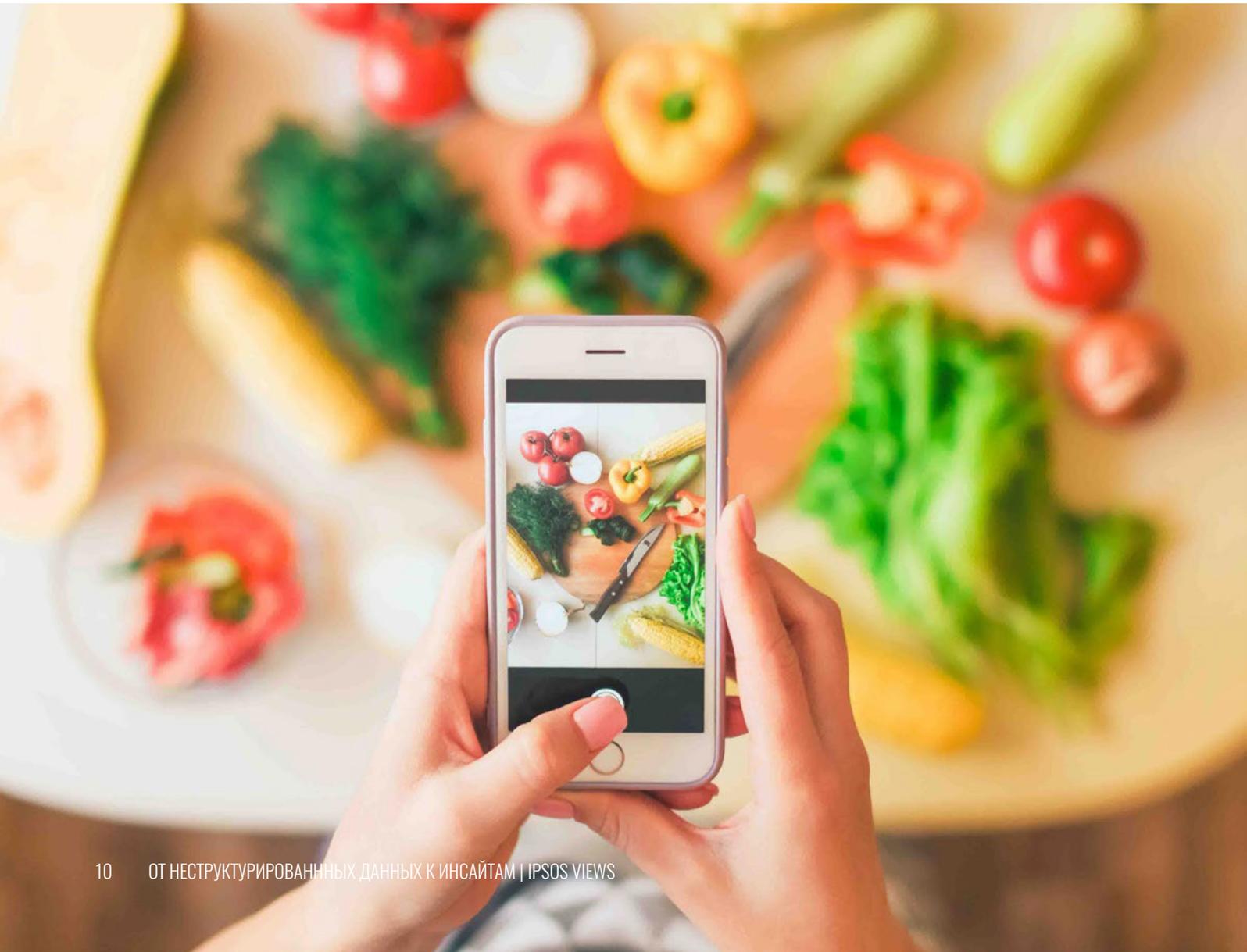
Просмотр данных человеком может также предотвратить ложные выводы. В начале пандемии мы стали отслеживать онлайн-обсуждения коронавируса, но так как наши аналитики оценивали эволюцию данных, мы заметили спад в упоминаниях.

В результате мы могли бы подумать, что важность этой темы упала. Напротив, коронавирус стал частью нашей повседневной жизни... Людям больше не нужно было упоминать его в своих постах. Используя возможности наших аналитиков, мы быстро выделили характеристики, отражающие многослойность опыта переживания пандемии потребителями – жизнь дома, развлечения, здоровье и благополучие и т.д. Чаще всего, рассмотрение только линии тренда не расскажет вам полную историю. Это показывает, как каждый слой работы с данными добавляет новую и важную информацию, чтобы помочь нам лучше понять, что происходит.

Рисунок 7. Обсуждения в соцсетях показывают, как эмоции менялись в отношении к еде



Источник: Synthesio, посты в социальных медиа, написанные на английском языке, март-апрель 2020



МИФЫ И РЕАЛЬНОСТЬ О СБОРЕ И АНАЛИЗЕ ДАННЫХ В СОЦИАЛЬНЫХ МЕДИА

Как и с любой другой растущей практикой, легко ошибиться в том, что возможно получить, и как это происходит. В этом случае легко принять за бриллиант подход, который окажется всего лишь обычным стеклом. Поэтому мы бы хотели развеять некоторые наиболее часто встречающиеся мифы.

Миф 1: Данные соцмедиа бесполезны – все эти данные из Twitter, они мне ничего не расскажут

Реальность: Данные Twitter в некоторых случаях могут преобладать, если специально их не ограничить. Иногда, это к лучшему. Даже микроблогинг с его короткими, но полезными постами имеет свою ценность, особенно его большой объем сообщений. Но данные Twitter годятся не для каждого вопроса или подхода. Так же, как вы бы задумывались, как собрать хорошую выборку в качественном или количественном исследовании, важно понимать свою генеральную совокупность и при сборе данных в социальных медиа. Прежде чем запускать исследования в соцмедиа, вы должны убедиться, что ваша платформа охватывает широкий круг источников⁴. В некоторых случаях специализированные блоги и форумы принесут больше пользы, чем Twitter. Также очень важно работать с данными, собранными из разных источников, как с гетерогенной массой. Например, твит о Tesla будет нести совершенно другое смысловое значение по сравнению с отзывом об опыте вождения Tesla или постом о Tesla на форуме о будущем автомобилей. Нужны опытные аналитики, чтобы реализовать такой проект. Точно также, как мы обычно разделяем данные опросов и в отчете сравниваем целевые группы, социоэкономические или отношенческие критерии.

Миф 2: Social intelligence нужен только для понимания наших действий по PR и маркетингу в соцмедиа, и чтобы вовремя решать проблемы негатива в соцсетях

Реальность: Да, данные социальных медиа и их анализ необходимы в таких случаях, но существует и множество дополнительных вопросов, с которыми они могут помочь. Их можно использовать, чтобы построить более точную сегментацию, чтобы понять ландшафт рынка и увидеть более широкий контекст, чтобы заметить сигналы макро-трендов и глубоко изучить микро- и нано-тренды; чтобы выявить, почему люди взаимодействуют (или нет) с определенными продуктами и услугами; чтобы понять специфичные бренду моменты, составляющие brand experience; и чтобы получить ясный ответ на вопрос, что побуждает людей вести или не вести себя определенным образом в ситуациях, связанных с самыми разными темами. И даже это – не исчерпывающий список, а только лишь те подходы, которые мы применяли для нахождения практических инсайтов.

Миф 3: Мы можем просто посмотреть, что интересного есть во всех данных социальных медиа

Реальность: Если у вас много времени и денег, можно поступить и так (потратив также огромное количество электроэнергии), но ресурсы редко бывают неограниченными. Как гласит мудрая истина: «Когда человек не знает, к какой пристани он держит путь, для него ни один ветер не будет попутным». Поэтому, важно задать некоторые стоп-сигналы для вашего исследования, чтобы понять, когда вы обнаружили что-то действительно интересное. Ваши экспертиза и знание индустрии помогут разработать соответствующие рамки.

Миф 4: Никто не говорит об этой теме, и поэтому я не могу провести исследование, используя social intelligence

Реальность: Иногда запросы могут быть очень специфичными. Это – не редкость в исследованиях брендов. Это не значит, что от изучения соцсетей не будет никакой пользы – на самом деле, они могут помочь нам глубже погрузиться в реальность потребителей.

Потребители могут не обсуждать определенный продукт бренда – например, они не говорят про замороженный горох или консервированные томаты – но они будут говорить об очень близких темах: как они готовят дома, любимые вкусы и блюда, и что они стремятся привнести в свой прием пищи.

Миф 5: Social intelligence – это быстро и дешево

Реальность: Анализ данных из соцсетей, определенно, может быть быстрее и дешевле применения традиционных методов. Но он также может показать более глубокую перспективу того, что движет людьми, потому что данные соцмедиа формируются без подсказок. Для подготовки качественной и пригодной для анализа базы данных, как и при интерпретации человеком для получения инсайтов, нужна многослойная доработка, поэтому хорошие сбор и анализ данных соцмедиа требуют определенного уровня затрат.

Миф 6: Все делает Искусственный интеллект

Реальность: Анализ неструктурированных данных был бы существенно ограничен без ИИ. В первую очередь, было бы практически невозможно найти данные, которые мы хотим изучить. Но только человек способен выделить смысл и распознать важность факторов, выделенных с помощью ИИ. Искусственный интеллект – это мощный инструмент, и он может показать нам паттерны, которые мы бы сами не заметили. Но нужен человек, чтобы сказать почему это важно. Конечно же, предполагается, что этот человек понимает вопрос и – более широко – категорию и тему.

Миф 7: Все инструменты одинаковы

Реальность: Каждый инструмент обладает своими плюсами и минусами, и его всегда можно заменить, в зависимости от того, что вы хотите сделать. Нужно учитывать множество факторов, в зависимости от того, что наиболее важно для вас и вашей организации – сбор данных, степень чистоты, простота использования, встроенный функционал, объем данных, уровень встроенных аналитики и ИИ, языковой и географический охват и так далее. Более того, эти факторы все время меняются. Их отслеживание может быть утомительным, но важно регулярно сверять ваши потребности с тем, что позволяют получать методы анализа.

Миф 8: Стать экспертом по изучению соцмедиа просто

Реальность: Как в любом исследовательском подходе, недостаток знаний может быть опасен. Нельзя быть экспертом во всех аспектах сбора и анализа данных соцмедиа. Но заинтересованный и любознательный ум может научиться выжимать максимум из неструктурированных данных и проводить надежное исследование соцмедиа с понятными и глубокими инсайтами.

Три составляющих элемента предполагают применение множества разнообразных навыков. Например, для поиска инсайтов подходят люди, которые обладают навыками количественных и качественных исследований, а также хорошим пониманием среды соцмедиа в целом и особенностей различных источников данных.

Мы также верим, что нашим клиентам необходимо знать больше, как использовать social intelligence и неструктурированные данные в правильном ключе.

Нашим клиентам важно правильно понимать основы, чтобы предотвратить 2 риска: попасться в ловушку «красивой упаковки» (и неверно применить подход) или попросту упустить возможности использования этого типа информации. Мы помогаем нашим клиентам ориентироваться в сложности мира сбора и анализа данных социальных сетей.

ОСНОВНЫЕ ВЫВОДЫ

Направление сбора и анализа данных социальных медиа только начинает развиваться и вызывает большой интерес. Как и при появлении маркетинговых исследований как дисциплины в целом, мы накапливаем знания, работая с кейсами, в которых можно выявить полезные смыслы и инсайты из баз данных.

Используя неструктурированные данные, особенно данные соцмедиа, нужно учитывать несколько важных факторов при оценке плюсов и минусов различных подходов:

- **Что мы действительно хотим узнать?** Чтобы не пытаться вычерпать ложкой море и не тратить лишние времена и деньги, вы должны понимать, ответы на какие вопросы вы ищете.
- **Говорят ли люди о предмет изучения?** Практичность и достижимость крайне важны. Любой хороший исследователь социальных медиа сделает проверку на эти критерии перед запуском проекта.
- **Какие источники наиболее важны для ответа на мой вопрос?** Существует множество разных источников. Люди пишут на форумах по другим причинам, чем они пишут в Twitter или Instagram. Каждый источник уникален, важно учитывать это, собирая данные, которые мы хотим исследовать.
- **Какие гипотезы и ожидания мы уже можем сформулировать?** Это хорошая практика для любого исследовательского проекта, но особенно для работы с наборами открытых данных. Если вы уже знаете, чего хотите, то лучше структурировать данные вокруг потребностей (например, есть

некоторые атрибуты, которые вы хотите отследить). В других случаях вы будете находиться в режиме поиска. Чем раньше вы определитесь, тем лучше будет подобран подход для достижения желаемого результата.

- **Как глубоко мы хотим посмотреть?** Как и в любом исследовательском подходе, можно изучить вопрос глубоко или поверхностно. Это применимо как к выполняемым с помощью ИИ, так и проводимым человеком составляющим проекта. Глубина даст больше понимания вопроса, но также, скорее всего, потребует больше времени и денег.
- **Как мы будем использовать данные?** Это более технический, но важный, вопрос, который связан с политиками защиты персональных данных. Об этом важно помнить. Разные компании опираются на разные подходы, какие данные отнести к публично доступным, а какие – нет. Удостоверьтесь, что политики компаний, с которыми вы работаете, не противоречат политикам вашей компании.

Эти факторы помогут вам провести успешную исследовательскую программу. Сбор и анализ данных социальных медиа могут очень многое предложить, как исследовательские инструменты. Они также могут стать очень полезным дополнением к другим источникам инсайтов, обеспечивая всестороннее понимание поведения и ожиданий потребителей. Продвигая подходы анализа информации из соцсетей с вниманием к деталям и с применением последних разработок ИИ, мы верим, что эти методы по-настоящему меняют природу исследований рынка.

ССЫЛКИ

1. <https://books.google.com/ngrams/info>
2. <https://www.ipsos.com/en/social-listening-tools-synthesioh>
3. <https://www.ipsos.com/en-us/knowledge/consumer-shopper/Ipsos-Pandemic-Adaptability-Continuum>
4. <https://www.ipsos.com/en/bigger-innovations-require-bigger-data>

ОТ НЕСТРУКТУРИРОВАННЫХ ДААННЫХ К ИНСАЙТАМ

Л. Де Воогд Глобальный руководитель Сервисной линии Social Intelligence Analytics, Ipsos

Т. Берд-Ноуланд Руководитель Сервисной линии Social Intelligence Analytics, Ipsos MORI

Серию статей
Ipsos Views выпускает
Ipsos Knowledge Centre.

www.ipsos.com
[@Ipsos](https://twitter.com/Ipsos)

GAME CHANGERS

