



Банк России



ИСПОЛЬЗОВАНИЕ БОЛЬШИХ ДАННЫХ В ФИНАНСОВОМ СЕКТОРЕ И РИСКИ ФИНАНСОВОЙ СТАБИЛЬНОСТИ

Доклад для общественных консультаций

Москва
2021

ОГЛАВЛЕНИЕ

Введение.....	2
1. Термины и определения.....	3
2. Тренды в использовании больших данных финансовыми организациями ведущих стран	5
2.1. Основные тренды в использовании больших данных	5
2.2. Области использования больших данных.....	9
3. Преимущества и риски использования больших данных.....	14
3.1. Преимущества использования больших данных	14
3.2. Риски использования больших данных.....	16
4. Использование больших данных в российских финансовых институтах	19
4.1. Направления использования больших данных в российских финансовых институтах	19
4.2. Организация процесса обработки и использования больших данных в российских финансовых организациях.....	21
4.3. Положительные эффекты от использования больших данных (по данным опроса).....	22
4.4. Риски использования больших данных российскими финансовыми институтами (по данным опроса).....	23
5. Подходы к регулированию рисков использования технологий больших данных за рубежом и возможные меры в России.....	25
5.1. Подходы к регулированию, применяемые за рубежом	25
5.2. Возможные шаги в области регулирования использования технологий больших данных в России в целях снижения рисков финансовой стабильности	28
Вопросы для консультаций	31

Материал подготовлен Департаментом финансовой стабильности. Ответы на вопросы, поставленные в докладе, а также замечания и предложения к нему просим направлять до 15 февраля 2022 года включительно по адресам: yudinats@cbr.ru; vorozhtsovpa@cbr.ru; mmm1@cbr.ru; burkovayua@cbr.ru.

Фото на обложке: Shutterstock/FOTODOM
107016, Москва, ул. Неглинная, 12
Официальный сайт Банка России: www.cbr.ru

© Центральный банк Российской Федерации, 2021

ВВЕДЕНИЕ

В последние годы становится все более распространенным применение технологий обработки больших данных в финансовом секторе. Данные технологии потенциально способствуют повышению качества предоставляемых услуг, помогают финансовым организациям снизить издержки и повысить эффективность деятельности.

Большие данные используются практически всеми крупными финансовыми организациями, а также консалтинговыми и технологическими компаниями, оказывающими услуги в финансовом секторе. Опрос Банка России, проведенный летом 2021 года, показал, что это характерно и для российского финансового рынка. При помощи таких данных финансовые институты могут в первую очередь более достоверно оценивать предпочтения потребителей финансовых услуг, делать продукты более персонализированными, корректнее оценивать риски, улучшать процедуры комплаенса и противодействия отмыванию доходов, полученных преступным путем, и финансированию терроризма (ПОД/ФТ).

Тем не менее использование больших данных может нести ряд рисков, в том числе имеющих системный характер для финансового рынка. Системные риски могут быть связаны с модельными рисками, способными привести к массовой некорректной оценке финансового положения заемщиков, рисками дискриминации потребителей (ценовой и неценовой), рисками нарушения конкуренции, рисками критической концентрации поставщиков данных (в том числе иностранных), рисками крупных утечек персональных данных. В связи с этим в последние годы в разных странах мира поднимаются вопросы регулирования применения финансовым сектором больших данных.

В настоящем докладе приводится описание трендов использования больших данных в мире и России, анализируются преимущества и риски их использования, а также возможные подходы к ограничению рисков применения больших данных в России, в том числе на основе практики регулирования, применяемой за рубежом. При этом настоящий доклад детально не затрагивает вопросы использования и защиты персональных данных.

1. ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ

Термин «*большие данные*» (*big data*, далее – *большие данные*) на текущий момент не имеет единого общепринятого определения. Этот термин используется для обозначения больших и/или сложных массивов данных и связанных технологий для их хранения и обработки. Указанные массивы данных могут иметь как структурированную форму (внешние и внутренние базы данных), так и неструктурированную (социальные сети, пресса и тому подобное).

Наиболее распространенным определением термина больших данных является следующее.

Большие данные – это информационный ресурс, удовлетворяющий критериям большого объема содержащейся информации, большой скорости ее обработки и большого разнообразия данных, который требует эффективных, экономичных и инновационных форм обработки информации и в результате позволяет улучшить анализ данных, принятие решений и автоматизацию процессов¹.

Таким образом, определение понятия больших данных предполагает наличие трех основных критериев: объем (*volume*), скорость (*velocity*) и разнообразие (*variety*), или 3-Vs², где:

- объем (*volume*) означает обширное количество данных, часто из различных источников;
- скорость (*velocity*) означает высокую скорость получения, сбора и анализа данных;
- разнообразие (*variety*) означает различные типы используемых данных, как структурированных, так и неструктурированных.

В соответствии с подходом UNECE³ можно [классифицировать различные виды больших данных следующим образом](#):

- информация из социальных сетей, блогов и интернет-сообщений;
- данные о деятельности в Интернете (в том числе поисковые запросы пользователей, данные о посещенных сайтах);
- информация традиционных бизнес-процессов (информация о транзакциях, покупках, заказах, платежах, регистрации клиентов, банковских операциях и тому подобном);
- данные государственных организаций (административные данные, в том числе таможенные, налоговые и другие, медицинские данные);
- данные мобильных и прочих устройств (данные геолокации, данные о трафике, данные систем типа «умный дом», камер видеонаблюдения, данные сенсоров, трекеров и тому подобного).

Поскольку обработка больших данных, как правило, происходит с применением технологий *искусственного интеллекта*, в этой связи используются следующие определения.

Искусственный интеллект – комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые как минимум с результатами интеллектуальной деятельности человека. Комплекс технологических решений включает в себя информационно-коммуникационную инфраструктуру, программное обеспечение (в том числе в котором используются методы машинного обучения), процессы и сервисы по обработке данных и поиску решений⁴.

¹ [Gartner IT-Glossary](#).

² Иногда отмечают дополнительные критерии, например точность/достоверность (*veracity*), ценность (*value*), волатильность (*volatility* – характеристика, указывающая на то, какое время данные являются полезными/релевантными). В этом случае концепция может носить название 4-Vs, 5-Vs и так далее.

³ *United Nations Economic Commission for Europe – Европейская экономическая комиссия ООН*.

⁴ Указ Президента Российской Федерации от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации» (вместе с «Национальной стратегией развития искусственного интеллекта на период до 2030 года»).

Машинное обучение – подкатегория искусственного интеллекта; способность компьютерных систем извлекать из данных знания и использовать их для дальнейшего решения задач⁵.

Технологии искусственного интеллекта – технологии, основанные на использовании искусственного интеллекта, включая компьютерное зрение, обработку естественного языка, распознавание и синтез речи, интеллектуальную поддержку принятия решений и перспективные методы искусственного интеллекта⁶.

Большие данные и искусственный интеллект тесно взаимосвязаны. Большие данные дают источник информации для анализа с использованием искусственного интеллекта. Технологии машинного обучения, основанные на нейросетевых подходах, позволяют зачастую получить наиболее полный и быстрый результат обработки больших данных. Однако здесь возникают определенные риски, связанные с прозрачностью и интерпретируемостью соответствующих моделей. При этом большие данные могут анализироваться не только с применением искусственного интеллекта. Для определенного круга задач эффективно могут применяться традиционные статистические (эконометрические) модели, однако именно искусственный интеллект позволяет извлечь максимальную пользу из работы с большими данными.

⁵ Определение сформулировано на основе различных источников, в том числе материалов доклада *Artificial intelligence and machine learning in financial services*, Financial Stability Board, 2017.

⁶ Федеральный закон от 24.04.2020 № 123-ФЗ «О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте Российской Федерации – городе федерального значения Москве и внесении изменений в статьи 6 и 10 Федерального закона «О персональных данных».

2. ТРЕНДЫ В ИСПОЛЬЗОВАНИИ БОЛЬШИХ ДАННЫХ ФИНАНСОВЫМИ ОРГАНИЗАЦИЯМИ ВЕДУЩИХ СТРАН

В последние годы финансовые институты все чаще используют в своей деятельности большие данные и методы их обработки. Однако наблюдаются различия как в масштабах использования больших данных среди различных сегментов финансового рынка, так и в подходах к их использованию. В данной главе проанализированы основные тренды в использовании больших данных зарубежными финансовыми институтами: масштабы использования больших данных в целом, характер взаимодействия со сторонними поставщиками и тенденции в прикладном использовании технологий больших данных для различных целей.

2.1. Основные тренды в использовании больших данных

1. Финансовый сектор все чаще использует технологии больших данных и технологии искусственного интеллекта для их обработки. Этому способствует увеличение объема информации в целом: домохозяйства и компании производят беспрецедентное количество данных, которые постоянно накапливаются; социальные сети позволяют отслеживать скорость распространения информации, компании подробно фиксируют процесс производства и продажи товаров и услуг; платежные транзакции и интернет-торговля также имеют свой «цифровой след». Растут инвестиции в большие данные компаний финансового сектора: по некоторым расчетам, в 2021 году они составят 14 млрд долларов США по сравнению с 9 млрд долларов США [в 2018 году](#).

Согласно исследованию **Европейского банковского управления (ЕБА)**, банки ЕС заинтересованы в использовании аналитики больших данных: растет доля как организаций, которые уже используют большие данные в своей деятельности (рост с 60 до 64% с 2018 по 2019 год), так и доля компаний, тестирующих, разрабатывающих и обсуждающих [внедрение новых технологий](#). Всего лишь 2% банковских институтов ответили, что они не занимаются внедрением аналитики больших данных.

2. Пандемия COVID-19, вероятно, способствовала ускорению внедрения технологии больших данных, а также машинного обучения и искусственного интеллекта для их обработки.

Обычно в период кризиса бизнес сокращает инвестиции в инновационные технологии

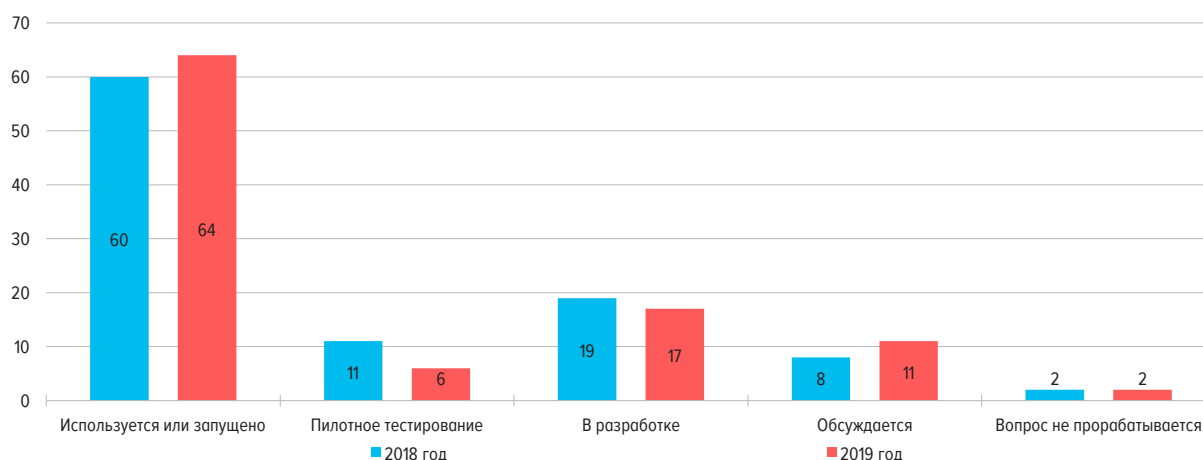
ИНВЕСТИЦИИ ФИНАНСОВОГО СЕКТОРА В БОЛЬШИЕ ДАННЫЕ В РАЗБИВКЕ ПО ТИПАМ ПРЕДОСТАВЛЯЕМЫХ ФИНАНСОВЫХ УСЛУГ (%)

Рис. 1



ИСПОЛЬЗОВАНИЕ АНАЛИТИКИ БОЛЬШИХ ДАННЫХ СРЕДИ БАНКОВ ЕС (%)

Рис. 2



Источник: [EBA](#).

и отдает приоритет удовлетворению краткосрочных потребностей в денежных средствах, а не долгосрочным технологическим проектам. Однако в августе 2020 года **Банк Англии** провел исследование использования банками машинного обучения, искусственного интеллекта и больших данных на фоне пандемии COVID-19¹, которое выявило совершенно иную тенденцию: общие инвестиции банков в технологии обработки больших данных возросли в период пандемии. Согласно исследованию, такая ситуация возникла из-за ряда причин:

- в целом возросла потребность в цифровизации бизнеса на фоне перехода потребителей на онлайн-каналы в условиях пандемии;
- желание финансовых институтов повысить эффективность через автоматизацию процессов и на фоне низких процентных ставок;
- сектор финансовых услуг в пандемию в целом пострадал не так сильно, как другие отрасли;
- увеличение количества данных на фоне более широкого использования населением во время пандемии компьютеров, смартфонов для торговли, удаленной работы и общения.

По мере того как банки подстраивают затраты и доходы в условиях пандемии COVID-19, они стремятся использовать машинное обучение и другие методы обработки значительных объемов данных для повышения эффективности и улучшения цифровых каналов доступа к клиентам. Пандемия простимулировала интерес к машинному обучению и другим методам обработки значительных объемов данных в банковском секторе Великобритании. Половина из опрошенных банков сообщила о росте важности технологий обработки больших данных для будущих операций и ни один из банков – о снижении. В части инвестиций равное количество респондентов (по 16%) заявили о снижении и увеличении объемов ресурсов, выделяемых на существующие программы обработки больших данных; при этом в отношении разработки новых программ 23% банков заявили о росте выделяемых ресурсов и только 11,5% банков – о снижении.

В качестве примера использования технологий больших данных в условиях пандемии COVID-19 можно привести создание индонезийским банком Bank Mandiri совместно с американским разработчиком Cloudera² специальной платформы больших данных Enterprise Information & Decision Platform, которая позволяет проводить мониторинг ликвидности и ежедневных операций филиалов в режиме реального времени с представлением наглядной карты

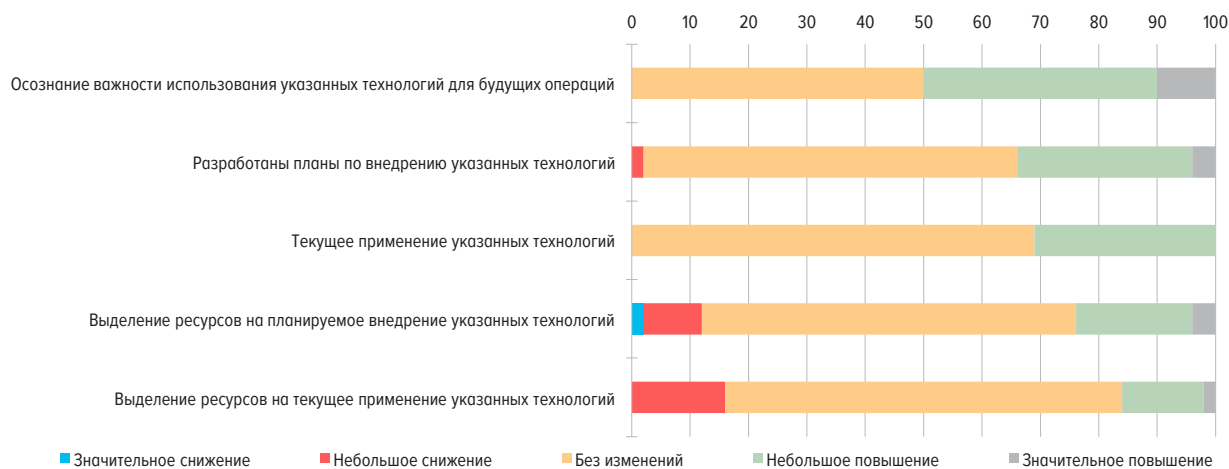
¹ [The impact of COVID on machine learning and data science in UK banking.](#)

² [Сотрудничество с Cloudera начало осуществляться еще в 2017 году.](#)

ВЛИЯНИЕ ПАНДЕМИИ COVID-19 НА ТЕКУЩЕЕ ИСПОЛЬЗОВАНИЕ И ПЛАНЫ ВНЕДРЕНИЯ МАШИННОГО ОБУЧЕНИЯ И ДРУГИХ МЕТОДОВ ОБРАБОТКИ ЗНАЧИТЕЛЬНЫХ ОБЪЕМОВ ДАННЫХ

Рис. 3

(%)

Источник: [The Bank of England](#).

зон воздействия COVID-19, на основании которой [принимается решение](#), какой филиал должен продолжать работать, а какой прекратить функционирование. Платформа также позволяет своевременно представлять соответствующую информацию клиентам и регулятору, анализировать факторы, которые могут ускорить процесс реструктуризации долгов клиентов, а также проводить мониторинг здоровья сотрудников банка на ежедневной основе, что позволяет минимизировать риск заражения.

3. В целом при работе с большими данными небольшие финансовые институты более склонны к использованию разработок сторонних поставщиков, тогда как крупные финансовые институты обладают достаточными ресурсами для разработки собственных решений.

Согласно результатам [исследования ЕВА](#), многие европейские финансовые институты используют разработки сторонних поставщиков (например, инструменты для разработки моделей); при этом ответственность за качество использованных моделей, созданных на основе разработок сторонних поставщиков, и их результатов возложена на сами финансовые институты.

В то же время в ходе [исследования Банка Англии и Службы финансового надзора Великобритании \(FCA\)](#) было выявлено, что британские финансовые компании в основном пользуются собственными программами для обработки данных посредством машинного обучения – 76% респондентов ответили, что программы для использования машинного обучения были разработаны и используются внутри самой компании. В то же время финансовые компании часто используют в процессе создания собственных программ для обработки данных элементы, разработанные сторонними поставщиками (модели, программное обеспечение и библиотеки машинного обучения), которые впоследствии адаптируются для внутреннего использования. Наиболее часто (36% случаев) разработками сторонних поставщиков пользуются небанковские финансовые организации. Возможно, это связано с небольшими размерами таких компаний и нехваткой ресурсов для собственных разработок.

Кроме того, финансовые институты все чаще используют услуги облачных сервисов для хранения данных, если наблюдается нехватка мощностей собственных серверов. Согласно [исследованию Банка Англии и FCA](#), 22% приложений финансовых институтов для обработки данных для машинного обучения работают на облачных серверах. Подчеркивается, что приложения, разработанные самим финансовым институтом, обычно работают на внутренних серверах. Облачные сервисы чаще используются небанковскими финансовыми институтами (они, как упоминалось выше, также чаще используют услуги сторонних поставщиков в части разработки самих приложений для обработки данных).

Примером использования услуг сторонних поставщиков технологий по обработке больших данных является заключение [соглашения между немецким Deutsche Bank и Google Cloud](#) для разработки высокотехнологических решений на основе использования машинного обучения и искусственного интеллекта, что в будущем поможет упростить банку анализ рисков, прогнозирование денежных потоков, выработку решений по управлению ценными бумагами клиентов. Кроме того, в рамках соглашения Google Cloud будет предоставлять немецкому банку услуги по облачному хранению больших данных. Один из крупных турецких банков – Akbank использует [американскую облачную платформу для обработки данных Datameer](#). [Платформа компании IBM](#) позволяет бразильскому банку Banco do Brasil обрабатывать большие данные в процессе риск-менеджмента.

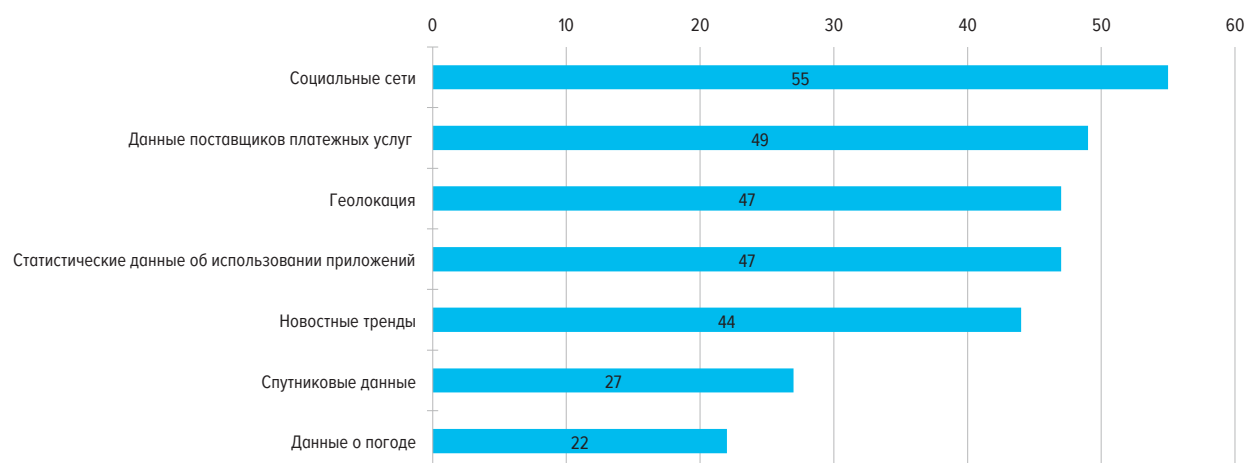
4. Помимо традиционных источников информации, финансовые институты используют альтернативные источники данных – от социальных сетей до данных спутников. [Такой подход](#) дает доступ к ранее не учитываемым данным, которые могут применяться для разных целей, включая торговлю ценными бумагами и инвестиции на основе данных (например, при принятии автоматизированных инвестиционных решений на фондовом рынке, при проведении кредитного скоринга, для совершенствования качества автоматизированных алгоритмов с помощью машинного обучения).

Согласно результатам [исследования EBA](#), для европейских банков собственные данные (информация о транзакциях клиентов, данные об использовании банковских продуктов, данные об обслуживании долга) являются основным источником для анализа. Внешние данные (данные социальных сетей, публичные данные о ценах на недвижимость, информация в СМИ и так далее) у некоторых банков вызывают опасения в части надежности и точности, поскольку аналитика с использованием данных низкого качества может привести к принятию неверных решений. По результатам [исследования Банка Англии и FCA](#), опрошенные финансовые компании используют внешние данные при применении технологий машинного обучения в 40% случаев, при этом применяются данные из различных источников (информация о потребителе для оценки кредитного риска, информация об автомобилях для страхования и так далее), которые могут анализироваться в сочетании с собственными данными для улучшения качества прогнозов.

В 2019 году Кембриджский университет и Всемирный экономический форум провели [исследование](#) о внедрении технологий искусственного интеллекта в финансовом секторе, в котором участвовала 151 компания из 33 стран (54% из них являются финтех-компаниями и 46% – традиционными финансовыми институтами). В работе, в частности, проанализированы основные источники внешних данных: наиболее часто используемым источником внешних

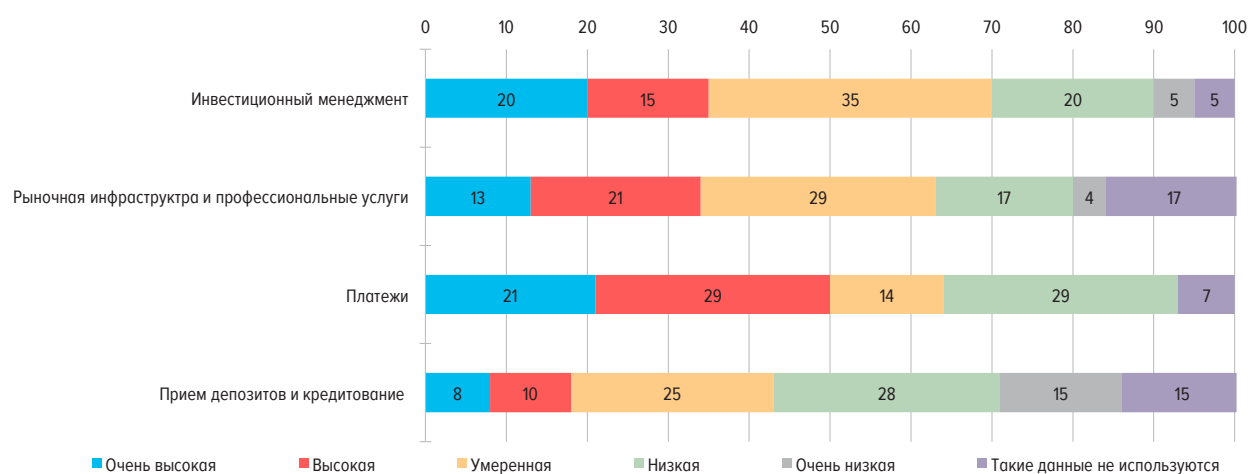
РАСПРОСТРАНЕННЫЕ ИСТОЧНИКИ ВНЕШНИХ ДАННЫХ (%)

Рис. 4



СТЕПЕНЬ ИСПОЛЬЗОВАНИЯ БОЛЬШИХ ДАННЫХ ПО ВИДАМ ФИНАНСОВЫХ ПРИЛОЖЕНИЙ (%)

Рис. 5



Источник: [Всемирный экономический форум](#).

данных являются социальные сети, данные платежных компаний и геолокации (их используют 55, 49 и 47% финансовых институтов соответственно). Наименее популярными источниками являются спутниковые данные и данные о погоде – такие данные обычно дороже, и для их обработки и интерпретации могут потребоваться специальные навыки.

Наиболее популярно использование внешних данных у платежных компаний (50% компаний данного сектора ответили, что в значительной степени учитывают внешние данные) и инвестиционных менеджеров (35% инвестиционных менеджеров масштабно пользуются внешними данными). Наименее популярно применение внешних данных у депозитных финансовых институтов (только 18% респондентов данного сегмента отметили [активное использование внешних данных](#)).

Важным источником больших данных о потребителях являются данные крупных высокотехнологичных компаний (бигтех). Эти данные затем используются связанными с ними финансовыми высокотехнологичными компаниями (финтех). Помимо повышения кастомизации финансовых услуг и улучшения возможностей по оценке заемщиков, это поднимает проблему конкуренции и необходимого государственного регулирования (см. подглаву 3.2.).

2.2. Области использования больших данных

1. Оценка кредитоспособности клиентов

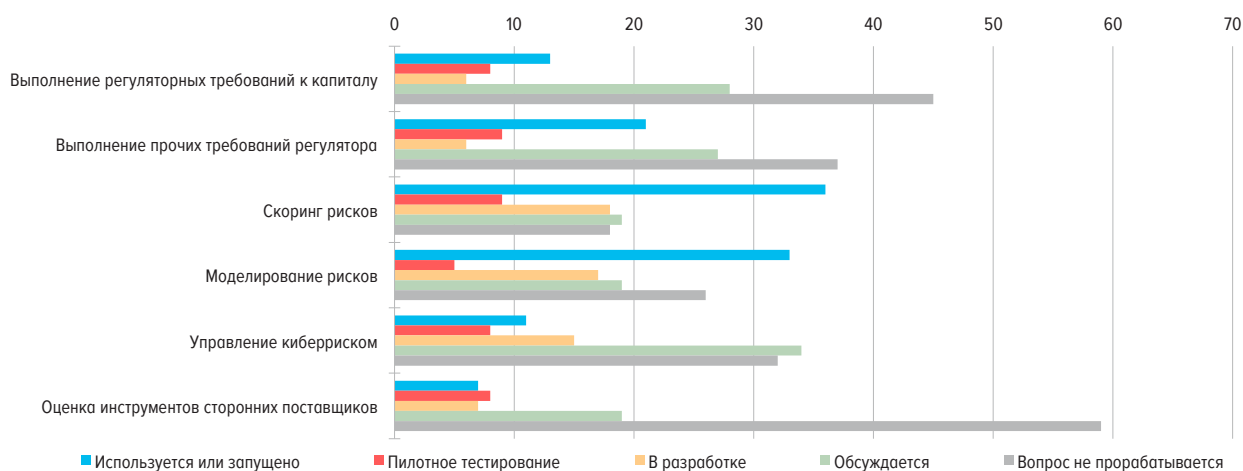
Оценка кредитоспособности клиентов – традиционная область для использования больших данных. Банки могут повысить качество кредитного скоринга новых клиентов с помощью сложных аналитических моделей с использованием значительных объемов данных, хранящихся у финансовых институтов (иногда в сочетании с внешними данными). Для доступа к платежной информации могут использоваться прикладные программные интерфейсы (application programming interface, API), что позволяет предлагать новые услуги для клиентов, например мгновенное кредитование или выдачу заранее одобренных кредитов новым клиентам.

[Исследование Банка Англии и FCA](#) выявило, что 57% опрошенных финансовых институтов используют технологии машинного обучения в своих моделях риск-менеджмента.

Как было сказано выше, большие данные, в том числе имеющиеся у бигтехов, все чаще используются для оценки кредитоспособности розничных клиентов. В Китае для оценки кредитоспособности заемщиков (в том числе малых и средних предприятий (МСП)) с использованием больших данных стали появляться компании, использующие данные об интернет-покупках:

ИСПОЛЬЗОВАНИЕ ЕВРОПЕЙСКИМИ БАНКАМИ БОЛЬШИХ ДАННЫХ ДЛЯ ЦЕЛЕЙ РИСК-МЕНЕДЖМЕНТА (%)

Рис. 6



Источник: [Всемирный экономический форум](#).

компания Alibaba запустила онлайн-банк MYbank, который использует данные из клиентской базы Alibaba. В 2015 году китайская компания Tencent запустила [интернет-банк WeBank](#), который для оценки кредитоспособности использует данные интернет-покупок, активности в социальных сетях и онлайн-играх.

В 2018 году пять крупнейших бразильских банков³ создали компанию [Quod – аналог бюро кредитных историй](#), которая при расчете кредитного риска заемщика учитывает наличие просроченных платежей, соблюдение графика погашений существующих займов. Проект использует аналитику больших данных. Технической стороной проекта занималась американская компания LexisNexis Risk Solutions.

[Чилийская финтех-компания Destacame запустила платформу кредитного скоринга](#), используя альтернативные данные (счета по коммунальным услугам и другую социально-экономическую информацию) для оценки кредитоспособности заемщиков в Чили и Мексике, которым банки обычно отказывают в предоставлении финансовых услуг. Компания сотрудничает с традиционными финансовыми институтами, предоставляя им альтернативную оценку кредитоспособности заемщика, и служит своеобразным посредником между финансовым институтом и заемщиком.

Стоит отметить, что иногда платформы больших данных для повышения доступности финансовых услуг создаются не частным сектором, а регуляторами. В 2020 году Монетарное управление Гонконга [объявило о намерении организовать новую инфраструктуру финансового рынка](#) – Commercial Data Interchange (CDI), которая должна повысить доступность финансовых услуг в Гонконге для МСП путем обеспечения безопасного и эффективного обмена информацией между банками и источниками коммерческих данных об операциях МСП. CDI будет предоставлять банкам проверенные данные о поведении потенциальных заемщиков (включая данные о платежах и тому подобном), которые банки смогут использовать для оценки кредитоспособности. Данные будут предоставляться банкам только [с согласия организаций](#).

2. Маркетинг и взаимодействие с клиентами

С помощью больших данных компании могут отслеживать различные аспекты поведения клиентов и получать сведения об их предпочтениях, а в дальнейшем на основании полученных данных направлять актуальные предложения по тем или иным продуктам банка, которые с большей вероятностью смогут заинтересовать клиента, так как будут соответствовать его

³ Banco do Brasil, Bradesco, Itaú, Santander и Caixa Econômica Federal.

индивидуальным потребностям. Финансовые институты могут использовать большие данные, полученные в результате различного взаимодействия с клиентами (в том числе обработки жалоб), для более точного определения и полного удовлетворения потребностей клиентов, в том числе путем использования чат-ботов. Кроме того, анализ продаж, продуктов и рынка в целом с применением больших данных может содействовать разработке новых финансовых продуктов, которые будут соответствовать потребностям клиентов и позволят добиться конкурентных преимуществ на рынке.

В своем [исследовании ЕВА](#) отмечает, что почти 95% опрошенных банков находятся в процессе проработки применения больших данных для взаимодействия с клиентами или уже используют (около 50% компаний) аналитику больших данных.

В банковской системе Франции цифровая трансформация охватывает в основном крупных игроков отрасли. Последние служат инкубатором для стартапов, функционирующих по модели *необанка*⁴. Например, в рамках одного из крупнейших французских конгломератов – [Société Générale](#) был создан необанк Prismae, который предлагает услуги цифрового банкинга для субъектов МСП и использует большие данные и искусственный интеллект для оказания индивидуальных консультационных услуг.

Британский банк Metro Bank⁵ использовал искусственный интеллект и прогнозную аналитику, чтобы помочь клиентам управлять своими финансами. Работая в партнерстве с Personetics, банк запустил службу Insights в приложении, которая отслеживает данные и шаблоны транзакций клиентов в режиме реального времени. Затем приложение предоставляет персонализированные подсказки для оплаты подписки и предупреждения о необычных расходах.

Канадская финансовая группа Canadian Western Bank в партнерстве с финтех-компанией Temenos Infinity предоставляет [персонализированные решения для субъектов МСП](#), позволяющие эффективно управлять своими денежными потоками и финансовыми решениями.

В Гонконге банк HSBC в 2018 году запустил [платформу Business Internet Banking](#), которая, анализируя данные по всем операциям на валютном рынке, совершаемым клиентами банка, дает подсказки и советы. Например, если клиент совершает несколько однотипных операций за определенное время, она может предложить автоматизировать процесс.

3. Управление активами

Инструменты анализа больших данных, в частности искусственный интеллект и машинное обучение, все чаще используются в управлении активами. Французские банки BNP Paribas и Société Générale работают над внедрением [искусственного интеллекта для анализа больших данных](#) с целью оптимизации доходности своих портфелей активов. Société Générale тестирует модели машинного обучения для формирования портфелей акций с целевым соотношением «доходность – риск». В модели используются большие данные, объединяющие 80 финансовых показателей (цена акции, EBITDA эмитента, коэффициенты рентабельности и так далее), и нефинансовые (ESG-характеристики, комментарии экспертов, информация СМИ).

BNP Paribas Asset Management, в свою очередь, внедрил технологии машинного обучения для скоринга эмитентов на основе анализа временных рядов их операционных и финансовых показателей, котировок акций, а также использует технологии обработки естественного языка (Natural language processing, NPL) и анализ настроений для оценки заявлений о деятельности компаний в СМИ.

Англо-немецкая компания Arabesque Asset Management разработала [платформу на основе больших данных](#) для оценки ESG-факторов более 7 тыс. крупнейших в мире компаний, котирующихся на бирже. На основе самообучающихся алгоритмов анализируются

⁴ Финтех-компании, предлагающие стандартные банковские услуги полностью в цифровом формате и не имеющие физических отделений.

⁵ [AI leaders in financial services Common traits of frontrunners in the artificial intelligence race.](#)

более 200 ESG-показателей из более 30 тыс. источников в 170 странах. Arabesque не только сама пользуется данной разработкой при формировании инвестиционного портфеля, но и предоставляет доступ к платформе другим организациям.

4. Страхование

Страховой сектор наряду с банками является лидером внедрения больших данных среди финансовых институтов. Многие страховые компании используют искусственный интеллект и большие данные для автоматизации процесса рассмотрения заявлений, выявления мошенничества и обеспечения соответствия регуляторным требованиям.

Кроме того, в автостраховании технология больших данных используется в связке с Интернетом вещей: основным источником данных для аналитики служат телематические устройства (черные ящики), позволяющие определить характер управления транспортным средством и, следовательно, точнее оценить риски. Данные, собираемые в реальном времени с использованием телематических устройств, позволяют страховым компаниям уточнять риск-профиль клиента и, соответственно, оптимизировать стоимость страховых услуг.

Страховые компании рассматривают большие данные в качестве основы своей будущей бизнес-модели и тарифной политики. Во Франции ряд страховых групп приступили к реализации «многоканальных стратегий»⁶ (*omnichannel strategies*) и использованию технологии больших данных и искусственного интеллекта для более эффективного анализа продаж, в том числе [составления прогноза объемов продаж страховых продуктов](#).

В феврале 2020 года Австралийская служба пруденциального регулирования (APRA) провела [опрос](#) 36 регулируемых страховых компаний по используемым новым технологиям в страховании (InsurTech), среди которых большие данные, искусственный интеллект, программное обеспечение как услуга, блокчейн, Интернет вещей (включая телематику). Согласно результатам опроса, большие данные – самая популярная категория InsurTech, используемая в основном для ценообразования и андеррайтинга.

5. Предотвращение реализации операционных рисков, в том числе киберрисков и рисков нарушения ПОД/ФТ

С помощью больших данных финансовые институты могут снижать издержки, а также повышать эффективность процедур управления операционным риском и систем комплаенса. Например, с помощью анализа большого объема данных можно упростить процесс выявления мошенничества. Важное направление применения анализа больших данных – обеспечение кибербезопасности (анализ киберугроз, выявление подозрительной деятельности, которая может нарушить работу внутренних систем банка).

В 2018 году банк HSBC запустил платформу Global Social Network Analytics для выявления и предотвращения финансовых преступлений с помощью использования больших данных. Платформа анализирует, в том числе в режиме реального времени, как собственные данные HSBC, так и внешние данные из различных источников (например, информацию о структуре владения компаниями или активами, которая позволяет выявить связи между контрагентами и транзакциями). Платформа анализирует транзакции с учетом 50 различных сценариев, пытаясь выявить какие-либо признаки отмывания денег, например скрытые связи между участниками сделки или подозрительную динамику платежей. Технологическое решение для анализа больших данных и внешние данные для платформы HSBC предоставляет [британская компания Quantexa](#). При этом [данные для Global Social Network Analytics хранятся на внутренних серверах HSBC](#).

⁶ Многоканальный маркетинг на основе больших данных предполагает сбор и унификацию большого массива данных для создания детальных профилей клиентов, их сегментацию и применение интеллектуальной бизнес-аналитики (*Business Intelligence*) для оптимизации маркетинговой стратегии.

6. Оптимизация отчетности и других процессов

Большие данные все чаще применяются для оптимизации и повышения операционной эффективности в процессе повседневной деятельности финансовых институтов, в частности для совершения операций, взаимодействия с регуляторами и повышения эффективности работы сотрудников. Некоторые финансовые институты, например Citibank, используют [большие данные для автоматизации процессов представления отчетности регуляторам](#). Немецкий Commerzbank применяет искусственный интеллект для автоматической [проверки сделок торгового финансирования](#) на соответствие законодательству перед непосредственным их совершением, а также [для составления отчетов](#) о прибылях и убытках. Еще один немецкий банк, Deutsche Bank, внедрил [различные технологии обработки больших данных](#) по торговым операциям и рискам.

Кроме того, большие данные могут использоваться для анализа возможностей повышения эффективности работы сотрудников. ЕВА, однако, [отмечает](#), что на данный момент указанное направление используется ограниченно.

В начале прошлого десятилетия сингапурский банк DBS начал использовать анализ больших данных для повышения эффективности управления персоналом и снижения текучести кадров. В рамках HR-департамента был создан аналитический отдел, который занимался оценкой трендов и определением риск-факторов текучести кадров. За восемь лет [показатель текучести кадров](#) снизился с 27 до 18% (ниже среднерыночного показателя).

С учетом вышеописанных трендов можно сказать, что в финансовом секторе интерес к технологиям, связанным с использованием больших данных, продолжит расти в будущем на фоне увеличения объема производимых данных и благодаря широким возможностям их использования в различных аспектах деятельности финансовых институтов.

3. ПРЕИМУЩЕСТВА И РИСКИ ИСПОЛЬЗОВАНИЯ БОЛЬШИХ ДАННЫХ

Большие данные могут использоваться в различных направлениях деятельности финансовых институтов, позволяя повышать эффективность, сокращать издержки и увеличивать прибыль. Технологии больших данных благодаря объемам обрабатываемой информации, скорости обработки данных и повышению уровня автоматизации дают финансовым институтам значительные преимущества в управлении рисками, помогают оптимизировать операционную деятельность и бизнес-процессы, усовершенствовать работу финансового рынка. Кроме того, большие данные дают возможность финансовым институтам предоставлять более персонализированные услуги потребителям и расширять объем клиентской базы. Тем не менее использование технологий больших данных может нести ряд рисков как для самих финансовых институтов, так и для потребителей финансовых услуг и финансовой системы в целом: методологические риски, риски в части надлежащей защиты персональных данных, риски, связанные с нарушением конкуренции и дискриминацией, риски сторонних поставщиков услуг и «серых» зон в регулировании.

3.1. Преимущества использования больших данных

Исходя из проанализированных выше направлений использования больших данных, можно сделать вывод о том, что большие данные вкупе с технологиями их обработки создают ряд важных преимуществ для финансовых организаций.

а) **Высокая скорость обработки значительных объемов информации** является одним из важнейших преимуществ технологий больших данных, которое помогает усовершенствовать различные аспекты деятельности финансовых институтов. Технологии больших данных позволяют, во-первых, быстро извлекать информацию из различных внешних источников (социальные сети, пресса и другие источники неструктурированных данных), а во-вторых, анализировать внешние или внутренние данные в режиме реального времени, в том числе с помощью использования машинного обучения. Высокая скорость обработки данных, [зачастую без временных лагов](#), позволяет усовершенствовать прогнозирование и анализ, а финансовые решения, принятие которых раньше занимало значительное время (денежные переводы, решения по выдаче кредитов, одобрение выдачи страховых полисов), выполняются гораздо быстрее, предоставляя потребителю почти [мгновенный доступ к финансовым продуктам](#).

На достаточно технологически развитом и широко использующем анализ данных (обычно структурированных – например, динамика цен на акции) фондовом рынке большие данные повышают скорость операций, в том числе за счет еще большей автоматизации и более широкого использования алгоритмов при принятии инвестиционных решений, прогнозировании будущих цен, для управления портфелем, при определении комиссий за услуги.

б) **Повышается качество риск-менеджмента** благодаря возможности проанализировать значительный объем данных (в том числе неструктурированных внешних данных, которые раньше не учитывались). Использование технологий искусственного интеллекта помогает выявить закономерности и взаимосвязи, например, в поведении потребителей, которые не могут быть выявлены при использовании традиционных инструментов. Использование больших данных в кредитном скоринге позволяет получить более качественную оценку рисков, в том числе при выдаче кредитов, способствуя более справедливому (с точки зрения риска) [формированию ставок](#). Кроме того, с помощью данных технологий можно повысить эффективность моделирования рисков, управления операционными рисками,

рисками нарушения ПОД/ФТ и киберрисками путем отслеживания поведения клиентов на предмет подозрительной активности с помощью искусственного интеллекта.

- с) **Повышение доступности финансовых услуг:** во многих развивающихся странах и странах с формирующимися рынками для домашних хозяйств и МСП доступность кредитов ограничена ввиду того, что банкам зачастую не хватает данных для оценки кредитоспособности потенциальных заемщиков. Например, [согласно некоторым оценкам](#), в Китае большинство жителей до сих пор имеют проблемы с доступом к кредитам, а на кредиты МСП приходится лишь 20–25% объема корпоративного кредитования, притом что вклад сектора в ВВП составляет примерно 60%, а вклад в занятость населения – порядка 80%. Использование больших данных может помочь устранить данную проблему, так как при оценке кредитоспособности заемщиков может использоваться не только информация о платежах по кредитам, но и информация о своевременности оплаты аренды жилья, коммунальных услуг, мобильной связи и так далее.

Большие данные могут помочь банковским финансовым институтам увеличить свою клиентскую базу за счет домохозяйств и компаний малого и среднего бизнеса, у которых нет кредитной истории, но кредитоспособность которых можно проанализировать с помощью альтернативных источников данных (для домохозяйств – арендные платежи, платежи за коммунальные услуги, для компаний – объемы продаж)¹, что повышает также доступность финансовых услуг в рамках [всей финансовой системы](#).

С другой стороны, в случае использования некорректных и неполных больших данных и/или некорректных моделей на их основе (модельный риск) доступность финансовых услуг для некоторых заемщиков может не увеличиться, а даже снизиться в результате неправильной оценки их кредитоспособности, как, например, показало исследование рынка потребительского кредитования США². Кроме того, доступность финансовых услуг может снизиться для некоторых категорий заемщиков в результате проявления дискриминации по тем или иным признакам с использованием моделей на основе больших данных (см. более подробно про модельный риск и риски дискриминации в подглаве 3.2.)

- d) **Возможность предоставления персонализированных услуг большему числу клиентов:** если раньше для выявления потребностей клиентов требовалось их личное общение с представителями финансового института, то сейчас анализ больших данных помогает предоставлять персонализированные услуги миллионам клиентов. Анализ трат потребителей, их поведения в социальных сетях и геолокаций потенциально дает возможность финансовым институтам лучше понять потребности клиентов. Кроме того, с помощью больших данных компании могут проводить более таргетированный маркетинг и ценообразование (особенно актуально для страхового сектора) благодаря более точной [сегментации потенциальных клиентов](#). Использование больших данных для более персонализированного подхода выгодно как клиентам на фоне повышения качества предоставляемых услуг, так и самим финансовым институтам (клиенты отдадут предпочтение компаниям с более качественными услугами).

Обработка больших массивов данных о клиентах помогает не только наилучшим образом удовлетворить потребности клиентов с помощью существующих продуктов, но и выявить, каких финансовых продуктов не хватает на финансовом рынке, что может быть использовано для формирования новых для клиентов продуктов. Создание новых продуктов на базе предварительного анализа потребностей клиентов с помощью больших данных положительно повлияет на клиентов (улучшение качества предоставляемых услуг), сами финансовые институты (повышение прибыли) и [развитие финансовой системы в целом](#).

¹ Тем не менее финансовым институтам следует учитывать возможное возникновение модельного риска.

² Big Data. A big disappointment for scoring consumer credit risk. National Consumer Law Center, 2014.

Таким образом, в целом технологии больших данных способствуют более качественной обработке информации в различных сегментах финансового рынка, что потенциально может снизить издержки и риски финансовых институтов и повысить эффективность и устойчивость финансового рынка в целом. С точки зрения финансовой стабильности использование больших данных позволяет снизить риски за счет улучшения качества риск-менеджмента и процессов работы с клиентами.

3.2. Риски использования больших данных

1. При использовании больших данных в риск-менеджменте и оптимизации операционной деятельности возникают **методологические риски, в том числе риски, связанные с качеством данных**. Методологии анализа больших данных пока находятся в процессе развития. До сих пор не ясно, как лучшим образом выбрать данные, их обработать и агрегировать. Также есть вопросы относительно аналитических инструментов, которые требуются для того, чтобы интегрировать результаты анализа больших данных с информацией, получаемой из традиционных источников.

Следует отметить, что использование внешних данных, связанных с распознаванием текста, анализом связей с помощью обработки неструктурированной информации из СМИ, социальных сетей и других источников, требует новых автоматизированных подходов к управлению качеством данных для обнаружения искажений фактов или дезинформации (fact checking).

Для повышения уровня полноты и качества больших внешних данных важно, чтобы используемые данные имели происхождение из независимых друг от друга источников.

2. Еще одной смежной проблемой является **модельный риск** при построении моделей на основе больших данных – использование ошибочных исходных данных или допущений, применение модели для целей иных, чем предполагалось при ее разработке, а также ошибки при разработке самой модели. Сложность вызывает то, что зачастую искусственный интеллект работает по принципу черного ящика (black box problem): данные закладываются в модель, искусственный интеллект проводит анализ и выдает результат, который невозможно перепроверить и интерпретировать³. В этой связи важен вопрос *квалификации сотрудников*, работающих с большими данными. На уровне исполнителей для минимизации ошибок и максимального использования возможностей больших данных специалисты должны понимать влияние исходных параметров, используемых в моделях, их взаимосвязи, а также оценивать потенциальное влияние параметров, не включенных в модель. Неполное понимание функционирования различных форм аналитики больших данных и качества данных или моделей может привести к реализации модельного риска. Высший менеджмент, в свою очередь, должен обладать достаточными знаниями для понимания [результатов моделирования и применимости моделей](#) к задачам, для которых они используются в финансовой организации.

Реализация модельного риска может привести к таким неблагоприятным последствиям, как систематическая некорректная оценка риска при использовании больших данных в риск-менеджменте. Может быть неправильно оценена платежеспособность клиентов при проведении кредитного скоринга, и данные ошибки могут носить не единичный, а массовый характер, что приведет к накоплению кредитных рисков в банковском секторе.

³ Нужно отметить, что в данном случае речь идет о моделях искусственного интеллекта, основанных главным образом на применении нейросетей. Однако большие данные могут анализироваться не только подобными методами, но и обычными статистическими (эконометрическими). И хотя в этом случае модельный риск также присутствует, модели являются гораздо более транспарентными и интерпретируемыми, так как в них, как правило, видны все факторы, влияющие на результат (отсутствует black box problem). Однако наиболее перспективным методом для работы с большими данными с точки зрения скорости и полноты получаемого результата считаются именно нейросетевые модели.

3. При использовании больших данных **для клиентов финансовых институтов возникают риски, связанные с защитой персональных данных**. Большие данные подразумевают сбор различной информации о клиентах. С одной стороны, это помогает повысить качество финансовых услуг за счет кастомизации. С другой стороны, увеличиваются риски ненадлежащего использования персональных данных и необходимость их защиты. Наиболее чувствительной является финансовая информация и информация о состоянии здоровья потребителей (которая может использоваться в первую очередь страховыми компаниями). Любое потенциальное ненадлежащее использование больших данных и недостаточная информационная безопасность могут подорвать доверие потребителей в долгосрочной перспективе. При этом зачастую люди сами делятся о себе значимой информацией, в том числе в социальных сетях, не до конца понимая важность той информации, которой они делятся, и спектра тех задач, для которых используется или может использоваться информация личного характера.
4. Большие данные позволяют более точно оценить потребности каждого потребителя индивидуально, но также позволяют финансовым институтам оценить готовность каждого конкретного клиента заплатить за услугу определенную стоимость, что потенциально может привести к возникновению **ценовой дискриминации**. То есть фактически одна и та же услуга может стоить по-разному для клиентов в зависимости от их платежеспособности или других обстоятельств. Данная ситуация приводит к снижению излишка потребителя, особенно если клиент нуждается в быстром предоставлении услуги (например, кредита или страхового полиса), он не провел заранее анализ цен на рынке или не имеет склонности к такому анализу, исходя из [истории его предыдущих покупок в Интернете](#). На основе анализа IP-адресов устройств, с которых покупатели выходят в Интернет, данных геолокации компании-продавцы могут отследить, живет ли покупатель в богатом или бедном районе, и в зависимости от этого предлагать разную цену.
5. Одним из следствий недостатка прозрачности и интерпретируемости методов обработки больших данных с использованием искусственного интеллекта являются **возможные проявления дискриминации (неценовой) по расовому, национальному, религиозному, гендерному и прочим признакам**. В силу недостаточной прозрачности указанных методов (в отличие от традиционных статистических моделей) дискриминация может быть неумышленной со стороны финансовой организации. Поскольку разным группам людей свойственны разные особенности поведения, в том числе в Интернете, социальных сетях, в процессе интернет-покупок и так далее, недостаточно качественная модель может на основе этих особенностей выдать результат, свидетельствующий о более низком уровне кредитоспособности заемщика по сравнению с его реальной кредитоспособностью в результате недоучета особенностей поведения разных групп населения. Таким образом, модели, основанные на больших данных, могут быть неточными и необъективными по отношению к [разным группам заемщиков](#).

При этом неценовая дискриминация теоретически может носить и умышленный характер со стороны финансовой организации опять же в силу недостаточной прозрачности используемых методов при работе с большими данными, в том числе для регуляторов и аудиторов.

6. **Неоднозначным является вопрос влияния использования больших данных на конкуренцию**. С одной стороны, использование больших данных крупными финансовыми институтами, которые изначально располагают значительным объемом информации о своих клиентах и обладают достаточными ресурсами для внедрения новых технологий, дает таким компаниям **конкурентные преимущества по сравнению с более мелкими игроками на рынке**. Большие базы данных позволяют наилучшим образом использовать элементы машинного обучения и искусственного интеллекта для скоринговой оценки заемщика, оценки риска страхователя или для предложения кастомизированной финансовой услуги, в то вре-

мя как отсутствие большого объема данных часто делает эти технологии [неэффективными в применении](#). С другой стороны, использование больших данных может стимулировать конкуренцию за счет выхода на рынок финансовых услуг финтех- и бигтех-компаний, которые зачастую используют технологии обработки больших данных, что в целом создает конкуренцию традиционным финансовым институтам.

7. Серьезными рисками для финансовой стабильности являются **риски сторонних поставщиков больших данных, услуг по их обработке и появление в связи с этим новых системных рисков**. Поскольку далеко не у всех финансовых организаций есть ресурсы и компетенции для работы с большими данными силами внутренних подразделений, вовлечение в эту деятельность новых, сторонних игроков (поставщиков внешней информации, разработчиков моделей анализа больших данных, облачных сервисов) приводит к усложнению взаимосвязей на финансовых рынках. Финансовые институты могут не иметь возможности в полной мере оценивать эти риски и управлять ими, так как данные риски выходят за рамки их организационной структуры. Масштабы новых взаимосвязей могут увеличить сложность финансовой системы и создать новые [каналы для распространения системных рисков](#). В частности, так как финансовые технологии, в том числе большие данные, постепенно стирают четкие функциональные границы между различными типами финансовых институтов, существует риск более быстрого распространения эффектов заражения между секторами и регионами. Концентрация на рынке сторонних поставщиков может привести к возникновению новых системно значимых игроков. Если доля нескольких крупных сторонних поставщиков на отдельных сегментах финансового рынка будет высока, то нарушения в операционной деятельности таких компаний могут привести к [широкомасштабным сбоям в других частях финансовой системы или экономике в целом](#).
8. Кроме того, вызывают опасения **риски, связанные с «серыми» зонами в регулировании**. Поставщики услуг в области работы с большими данными или новые типы компаний, в том числе поставщики больших данных и моделей на их основе, могут выпасть из полноценного регуляторного периметра финансовых регуляторов⁴. В результате могут появляться «серые» области финансовой деятельности, [не подлежащие регулированию](#). Особую опасность данная ситуация будет представлять в случае, если компании, выходящие за пределы регуляторного периметра, станут системно значимыми игроками на рынке.
9. Еще одним вызовом является **повышение ряда рисков для финансовых институтов, заставляющих с внедрением технологий больших данных**. Ряд финансовых институтов успешно используют большие данные для мониторинга и предотвращения реализации операционных рисков, в том числе киберрисков, и рисков нарушения ПОД/ФТ. Ожидается, что в будущем все больше финансовых институтов будут внедрять данные технологии. Однако существует риск того, что мошенники обратят внимание на финансовые институты, которые отстают во внедрении данных технологий (зачастую речь идет о малых и средних финансовых институтах, у которых может не хватать ресурсов для внедрения новых технологий). *Таким образом, использование технологий больших данных финансовыми институтами имеет значительные преимущества, однако также несет в себе существенные риски. Стоит отметить, что часть рисков связана с недостаточным опытом использования больших данных финансовыми институтами (например, нехватка квалифицированных кадров, методологические риски), и по мере развития рынка данные риски, возможно, будут устраняться. Для извлечения максимальной пользы от технологий больших данных регуляторы финансового рынка прорабатывают рекомендации и требования, направленные на развитие культуры этичного использования персональных данных, ответственного применения искусственного интеллекта и аналитики данных (см. главу 5 настоящего доклада).*

⁴ При этом операторы персональных данных в РФ в любом случае подпадают под компетенцию Роскомнадзора.

4. ИСПОЛЬЗОВАНИЕ БОЛЬШИХ ДАННЫХ В РОССИЙСКИХ ФИНАНСОВЫХ ИНСТИТУТАХ

Банк России летом 2021 года провел опрос крупнейших российских финансовых институтов на предмет использования больших данных в их деятельности. По результатам опроса Банком России были проанализированы основные тенденции и связанные с ними риски. В опросе приняли участие 13 банков, включая 9 СЗКО, 2 БКИ, 1 биржа, 1 крупная страховая компания, 1 платежная система, 1 система денежных платежей и переводов.

4.1. Направления использования больших данных в российских финансовых институтах

По мнению подавляющего большинства организаций, принявших участие в опросе, в последние 3–5 лет в России наблюдается **рост использования больших данных** и интереса к ним в целом. При этом, по мнению большинства участников опроса (58%), **пандемия не оказала заметного влияния на цифровую трансформацию бизнес-модели**, по мнению же остальных (42%), пандемия простимулировала использование технологий больших данных.

Почти все кредитные организации и подавляющее большинство некредитных финансовых организаций, принявших участие в опросе, включили развитие использования технологий обработки больших данных в **стратегию своей деятельности**.

Все принявшие участие в опросе **кредитные организации** сообщили, что уже достаточно широко и успешно используют в своей деятельности большие данные и планируют дальнейшее расширение их использования в большинстве своих бизнес-процессов.

Основное направление использования – кредитный скоринг. Большие данные используются при разработке моделей оценки риска как розничного, так и корпоративного кредитования. С помощью технологий больших данных происходит формирование предодобренных предложений для клиента, совершенствуются процессы сбора просроченной задолженности по кредитам.

Большие данные используются в риск-моделировании, поведенческом анализе клиентов, а также для целей информационной безопасности и противодействия мошенническим действиям, направленным против клиентов банка (антифрод-системы по выявлению атипичного поведения клиентов и другие), в автоматизации ответов на запросы клиентов (чат-боты).

При помощи больших данных вырабатываются наиболее активные каналы коммуникации с клиентом (CRM), разрабатываются модели кросс-продаж, осуществляется оценка тарифов для клиентов.

Кроме того, в рамках работы с большими данными некоторые банки осуществляют анализ транзакций клиентов для определения (корректировки) кредитных рейтингов корпоративных заемщиков и ИП.

В рамках мониторинга рисков контрагентов – финансовых организаций осуществляется анализ новостного фона с целью своевременного выявления негативных сигналов о финансовой устойчивости контрагентов. При помощи новостной аналитики выявляются взаимосвязи между заемщиками. Здесь, помимо методов анализа данных на основе машинного обучения, все более актуальными становятся методы и средства анализа графов, например, для исследования аффилированности участников финансового рынка, для анализа денежных потоков.

Некоторые банки с использованием больших данных строят модели PD/LGD заемщиков, а также рассчитывают величину экономического капитала.

Большинство опрошенных банков отмечают, что от 90 до 100% розничных кредитных заявок обрабатываются с использованием технологий больших данных, в целом по портфелю активов доля использования больших данных составляет в разных банках от 30 до 90%.

Некоторые банки используют роботов при формировании портфелей ценных бумаг, в том числе роботов-советников для клиентов. Роботы перебирают все возможные стратегии и выбирают наилучшую.

Кроме того, большие данные используются для целей автоматизации внутренних процессов банков.

По степени внедрения технологий больших данных можно отметить, что **среди различных направлений деятельности кредитных организаций уровень продвинутой работы по внедрению указанных технологий не одинаков**. Так, в риск-менеджменте, работе с клиентами, а также андеррайтинге и ценообразовании большинство респондентов отмечают масштабное внедрение соответствующих технологий или внедрение в достаточном объеме. При этом в анализе рынка, трейдинге, управлении активами, а также проведении платежей в большинстве банков лишь прорабатывается вопрос внедрения технологий на основе больших данных (таблица 1).

Среди направлений риск-менеджмента банки отмечают использование больших данных преимущественно в управлении кредитными рисками, в особенности розничными (таблица 2). При этом оценка других типов риска (рыночного, ликвидности и других), а также оценка достаточности капитала происходят с использованием больших данных пока еще в меньшем количестве банков. Так, в некоторых уже внедрено, в других только обсуждается, а в части банков нет планов по внедрению (таблица 2).

Страховые компании используют анализ больших данных в андеррайтинге, а также для выявления спроса на страховые продукты (клиентская аналитика), анализ больших данных по убыточности и аварийности для формирования тарифных правил. **В страховых компаниях** доля использования больших данных при формировании розничного портфеля достигает 100%, в целом по портфелю – 60%.

Биржи используют большие данные для целей комплаенса, анализа клиентского поведения, а также для разработки и предоставления аналитических продуктов клиентам.

Операторы платежных систем используют большие данные для планирования маркетинговых акций, анализа рынка, риск-менеджмента и проведения платежей.

ИСПОЛЬЗОВАНИЕ ТЕХНОЛОГИЙ НА ОСНОВЕ БОЛЬШИХ ДАННЫХ В РАЗЛИЧНЫХ НАПРАВЛЕНИЯХ ДЕЯТЕЛЬНОСТИ ФИНАНСОВЫХ ОРГАНИЗАЦИЙ И СТЕПЕНЬ ИХ ВНЕДРЕНИЯ (ПО ДАННЫМ ОПРОСА) Табл. 1

	Не используется и пока не планируется (%)	Рассматривается вопрос о возможном внедрении (%)	Фаза разработки (%)	Ограниченное внедрение (%)*	Внедрено в достаточном объеме (%)*	Масштабное внедрение (%)*
Риск-менеджмент	0	8	0	17	42	33
Взаимодействие с клиентами (работа колл-центров, создание персонализированных предложений и т.п.)	0	0	0	33	42	25
Анализ рынка	17	42	8	17	17	0
Андеррайтинг и ценообразование	0	17	25	8	25	25
Трейдинг	33	17	17	8	17	8
Управление активами	17	33	17	8	17	8
Проведение платежей	25	33	17	8	17	0
Другие процессы операционной деятельности, в т. ч. ПОД/ФТ	10	10	0	80	0	0
Другое	0	0	67	33	0	0

СТЕПЕНЬ ВНЕДРЕНИЯ ТЕХНОЛОГИЙ БОЛЬШИХ ДАННЫХ В СФЕРЕ РИСК-МЕНЕДЖМЕНТА В ФИНАНСОВЫХ ОРГАНИЗАЦИЯХ (ПО ДАННЫМ ОПРОСА)

Табл. 2

	Пока нет планов по использованию (%)	Обсуждается возможное внедрение (%)	Фаза разработки (%)	Пилотное тестирование (%)	Используется (%)
Оценка достаточности капитала, в том числе:	17	33	25	0	25
при применении ПВР	18	36	27	0	18
при применении ВПОДК	27	18	27	0	27
Оценка активов, в т.ч. по МСФО	9	27	27	9	27
Выполнение регуляторных требований, а именно:					
требования к капиталу	55	9	9	0	27
комплаенс, в т.ч. требования по ПОД/ФТ	18	18	18	9	36
Другое					
Оценка рисков, в том числе:	10	0	0	10	80
розничное кредитование	0	9	9	0	82
кредитование МСП	9	9	9	18	55
корпоративное кредитование	18	0	9	18	55
оценка рыночного риска по торговому портфелю	36	18	18	0	27
оценка операционного риска	36	18	9	9	27
оценка риска ликвидности	36	18	9	9	27
оценка процентного риска по банковскому портфелю (ALM)	27	18	9	9	36
прочие риски	0	20	20	0	60
Управление киберрисками	40	20	10	0	30

БКИ используют большие данные для создания скоринговых продуктов, анализа рынка и разработки продуктов.

В целом большинство организаций (58%) отмечают использование больших данных в широком спектре своих бизнес-процессов, и лишь меньшая часть респондентов отмечает их внедрение только в одном или нескольких бизнес-процессах (33%) или наличие пилотной стадии применения у себя соответствующих технологий (8%).

4.2. Организация процесса обработки и использования больших данных в российских финансовых организациях

Процесс обработки и хранения больших данных организован в подавляющем большинстве финансовых организаций следующим образом. Создано единое хранилище данных, за работу и развитие которого отвечает отдельное подразделение (департамент по работе с данными и тому подобное), и, кроме того, в каждом бизнес-блоке имеются свои специалисты или подразделения, ответственные за работу с данными. При этом в некоторых финансовых организациях указанное подразделение на уровне организации несет ответственность также за разработку всех систем и моделей, связанных с использованием искусственного интеллекта, однако в большинстве финансовых организаций за разработку моделей отвечают все-таки собственные функциональные специалисты (подразделения) в каждом бизнес-блоке.

Помимо внутренних источников больших данных, финансовые организации собирают большой набор **внешних данных**. В первую очередь используются следующие источники (в порядке убывания частоты упоминания):

83%	данные операторов сотовой связи
67%	платежные данные
58%	данные геолокации
33%	данные социальных сетей
33%	новости из прессы
25%	спутниковые данные
17%	другое
17%	используются только внутренние источники данных

Все финансовые организации отметили потребность в дополнительных **данных государственных органов**. В частности, банки хотели бы получить данные Госуслуг, ФНС (данные о налогах и транзакциях), ПФР, МВД (уголовные и административные дела), ФССП (данные об исполнительных производствах), ФМС, Росреестра (включая сведения о движимом и недвижимом имуществе в собственности), ЕГРЮЛ, ЕГРИП, системы ЖКХ (платежи), ЗАГС (информация о семейном положении), ГИБДД, Минтруда (данные по занятости), судов (иски на рассмотрении), Банка России (в том числе по платежным операциям, СБП), паспортные данные и телефоны граждан, адреса регистрации, номера СНИЛС и ОМС, данные о состоянии здоровья граждан. По юридическим лицам банки также интересуют полные данные о привлеченных средствах во всех кредитных организациях. Указанные данные, по мнению финансовых организаций, позволят им более корректно оценивать финансовое положение заемщиков, предлагать им более качественные продукты и услуги.

Большинство финансовых организаций отметили, что **обезличенные микроданные** представляют потенциальную ценность, но на текущий момент не используются. В частности, банки хотели бы получить обезличенные микроданные от БКИ, которые могли бы быть использованы в моделях прогноза поведения клиентов в зависимости от макроэкономической ситуации. Один из банков сообщил, что планирует использовать обезличенные микроданные операторов связи для прогнозирования клиентских потоков в офисах банка. Одна из причин ограниченного использования обезличенных данных, по мнению финансовых организаций, – регуляторные пробелы, в частности, необходимо более четко определить понятие «обезличенные данные» и определить их статус правообладания. Некоторые финансовые организации отмечали, что в настоящий момент подходы к обезличиванию данных регуляторно не определены, в результате передача детальных данных клиентов невозможна без существенной потери качества данных, также отсутствует единый стандарт передачи данных.

Однако ситуация с использованием обезличенных микроданных может измениться в будущем: в ноябре 2021 года Президент Российской Федерации Владимир Путин поручил принять законодательные решения, которые обеспечат доступ российских разработчиков искусственного интеллекта, научных организаций, бизнеса к [массивам обезличенных данных государства](#).

4.3. Положительные эффекты от использования больших данных (по данным опроса)

По мнению финансовых организаций, принявших участие в опросе, использование больших данных имеет безусловно положительный экономический эффект, но его конкретный денежный эквивалент оценить затруднительно, поскольку технологии обработки больших данных уже являются неотъемлемой частью многих бизнес-процессов и их финансовый результат не выделяется отдельно. Эффект также выражается в качественном улучшении бизнес-процес-

сов. Один из опрошенных банков оценил эффект использования больших данных на уровень прибыли в размере порядка 20%. Биржа оценила ежегодный рост выручки благодаря использованию больших данных на 3% и более. При этом большинство организаций отметили, что видят рост потенциала получения прибыли от использования больших данных в дальнейшем.

Среди **основных преимуществ, которые дает использование больших данных** для финансовой системы (как на уровне организаций, так и на макроуровне), российские финансовые организации отметили следующие (в порядке убывания частоты упоминания):

- повышение качества риск-менеджмента (100% респондентов);
- оптимизация операционной деятельности и бизнес-процессов (100%);
- возможность предоставления персонализированных услуг большему числу клиентов (100%);
- увеличение клиентской базы (75%);
- помощь в разработке новых продуктов (67%);
- повышение доступности финансовых услуг (67%);
- помощь в решении задач ПОД/ФТ, а также информационной и экономической безопасности (58%).

4.4. Риски использования больших данных российскими финансовыми институтами (по данным опроса)

Среди рисков использования больших данных **банками** были отмечены в первую очередь следующие:

- методологические и модельные риски, в том числе риски, связанные с качеством и полнотой данных, и квалификацией сотрудников, работающих с моделями (92%);
- риски, связанные с защитой персональных данных (67%);
- рост зависимости от сторонних поставщиков услуг (облачных решений, разработчиков платформ и моделей) (58%);
- риски, связанные с «серыми» зонами в регулировании (42%);
- риски, связанные с возникновением ценовой дискриминации (25%).

Другие финансовые институты (помимо банков), принявшие участие в опросе, отметили в первую очередь методологические и модельные риски и риски, связанные с защитой персональных данных, а также рост зависимости от сторонних поставщиков услуг.

Как показал опрос, с целью снижения рисков, связанных с защитой персональных данных, в большинстве банков (75%) и других финансовых организаций хранение больших данных происходит только на собственных серверах, однако некоторые банки, помимо собственных, используют также внешние (облачные) хранилища данных. Те организации, которые используют внешние хранилища данных, применяют специальные меры для обеспечения сохранности и конфиденциальности данных, предусмотренные законодательством и внутренними стандартами, в том числе обезличивание данных, шифрование, многофакторную аутентификацию при доступе к данным и другие. Соблюдение данных мер контролируют и проверяют независимые подразделения информационной безопасности.

С целью снижения рисков, связанных с «серыми» зонами в регулировании, Банком России в настоящий момент разрабатывается ряд законодательных инициатив. Так, прорабатываются вопросы, касающиеся особенностей оборота данных, находящихся в распоряжении финансовых организаций. Также совместно с Минцифры России разрабатывается регулирование оборота обезличенных персональных данных.

С целью повышения качества управления методологическими и модельными рисками в финансовых организациях, рисками зависимости от сторонних поставщиков данных и услуг по их обработке, а также рисками различных форм дискриминации в настоящем докладе предлагается ряд шагов в области развития принципов управления указанными рисками (см. главу 5 настоящего доклада).

Таким образом, по итогам опроса, проведенного Банком России среди крупнейших финансовых организаций, было выявлено, что технологии работы с большими данными уже играют значительную роль в деятельности финансовых организаций и эта роль будет расти. Указанные технологии применяются в большинстве ключевых бизнес-процессов и положительно влияют на финансовый результат. Основные направления применения технологий больших данных – кредитный скоринг и риск-менеджмент, а также клиентская аналитика. Финансовые организации хотели бы существенно расширить набор внешних источников больших данных за счет баз данных органов государственной власти. При этом существуют риски использования больших данных, в том числе персональных данных, для ограничения которых банки применяют процедуры по защите данных. В большинстве финансовых организаций создано специальное централизованное подразделение по работе с большими данными.

5. ПОДХОДЫ К РЕГУЛИРОВАНИЮ РИСКОВ ИСПОЛЬЗОВАНИЯ ТЕХНОЛОГИЙ БОЛЬШИХ ДАННЫХ ЗА РУБЕЖОМ И ВОЗМОЖНЫЕ МЕРЫ В РОССИИ

На финансовом рынке существует большой интерес к использованию технологий обработки больших данных, так как они предоставляют финансовым институтам возможность лучше оценивать свои риски и сокращать расходы, а клиентам – получить дополнительные выгоды. Тем не менее использование данных технологий сопровождается рисками, и задачей регулятора должно быть их ограничение.

5.1. Подходы к регулированию, применяемые за рубежом

Регуляторные режимы, применяемые к технологиям обработки больших данных, в том числе к искусственному интеллекту, находятся на достаточно ранних этапах разработки. Например, согласно [исследованию ЕВА](#), текущее регулирование на данном этапе в достаточной степени покрывает вопросы, связанные с использованием информационных технологий (в том числе риски, связанные с защитой данных, качеством данных; обеспечение непрерывности и устойчивости деятельности), безопасностью и управлением. Дальнейшее развитие регулирования может быть сконцентрировано на вопросах управления данными и этических аспектах.

На данный момент требования и руководства регуляторов в части использования технологий обработки больших данных зачастую являются комплексными и охватывают следующие вопросы: **обеспечение недискриминации пользователей, этическое применение технологий больших данных (в том числе в части защиты персональных данных) и прозрачность в интерпретации результатов моделей.**

Например, в январе 2019 года **Департамент финансовых услуг штата Нью-Йорк (NYDFS)** опубликовал [циркуляр](#) с руководящими принципами (требованиями) по использованию альтернативных данных страховщиками жизни. Принципы закрепляют, что использование таких данных не должно приводить к возникновению незаконной дискриминации потребителей; страховые компании должны удостовериться, что алгоритмы и модели соответствуют принципам проведения актуарной оценки с достаточным обоснованием любой заявленной корреляции или причинно-следственной связи. Кроме того, страховая компания должна раскрывать клиентам содержание и источник любых внешних данных, на основании которых страховщик принял неблагоприятное решение.

В июле 2021 года **Комиссия по финансовым услугам Республики Корея (FSC)** опубликовала [руководство](#) с минимальными требованиями к финансовым институтам в части использования искусственного интеллекта: требования касаются внедрения необходимых механизмов внутреннего контроля, обеспечения надлежащего использования персональных финансовых данных, предотвращения дискриминации, а также обеспечения защиты прав потребителей. Требования станут обязательными после проведения необходимой подготовительной работы (дата не уточняется).

С 2018 года в ЕС вступил в силу [Общий регламент о защите персональных данных физических лиц](#) (General Data Protection Regulation, GDPR), в котором закреплены основные принципы работы с персональными данными, которые необходимо учитывать также при анализе регулирования финансовых данных. В частности, принципы, заложенные в GDPR, предусматривают, что обработка персональных данных обеспечивается получением целевого, явно выраженного согласия на их использование, а обеспечение безопасности данных начинается с проектирования систем (privacy by design). За несоблюдение требований GDPR устанавливаются существенные штрафы от оборота компаний (до 20 млн евро или 4% от годового оборота компании в зависимости от того, какая сумма будет выше). Данный регламент значительно повлиял на работу с большими персонализированными массивами данных и аналитику на их основе.

Ряд регуляторов также опубликовал рекомендации для финансовых институтов, направленные на обеспечение недискриминационного и этичного использования технологий больших данных, а также прозрачности моделей. **Европейская служба по надзору в сфере страхования и пенсионного обеспечения** (EIOPA) опубликовала [принципы](#) управления использованием искусственного интеллекта в страховом секторе. Документ содержит в том числе следующие рекомендации: соблюдение принципа справедливости и недискриминации при применении технологий, связанных с искусственным интеллектом, принципа прозрачности и объяснимости используемых моделей, необходимость надзора со стороны сотрудников на протяжении всех этапов использования искусственного интеллекта. Подчеркивается необходимость соблюдения принципов управления данными, обеспечения устойчивости работы инфраструктуры и систем, связанных с искусственным интеллектом.

В ноябре 2021 года **ЕВА** опубликовало [консультативный доклад](#) о принципах надлежащего использования технологий машинного обучения в рамках моделей внутренних рейтингов (IRB) с целью расчета регуляторного капитала для покрытия кредитного риска. В частности, документ включает следующие принципы: необходимость соответствующего уровня квалификации персонала для разработки и проверки моделей машинного обучения; предоставление высшему руководству документации о функционировании модели машинного обучения. Рекомендуется избегать излишних усложнений модели. Кроме того, модель должна быть пригодна для корректной интерпретации, следует также обеспечить наличие объясняемого результата и надежную валидацию модели.

Федеральное управление по финансовому надзору Германии (BaFin) опубликовало надзорные принципы ответственного использования больших данных и искусственного интеллекта [финансовыми институтами](#). Данные принципы представляют предварительные идеи для последующей разработки минимальных надзорных требований. В документе представлены общие принципы использования алгоритмов при принятии решений (необходимость разграничить обязанности внутри организации, разработки системы управления рисками, адаптированной под использование алгоритмов при принятии решений; обеспечение отсутствия смещенности в результатах при принятии решения на основе алгоритмов и дискриминации клиентов). Кроме того, рассмотрены специфические принципы, которые необходимо учитывать при внедрении алгоритмов, основанных на больших данных и искусственном интеллекте (необходимость разработки стратегии проверки качества данных в алгоритмах и соблюдения требований к защите персональных данных; разработка документации по проведению верификации алгоритмов как самостоятельно компанией, так и аудиторами и органами надзора; эффективные процедуры валидации алгоритмов и выборки данных), и специфические принципы, которые необходимо учитывать при применении и интерпретации алгоритмов (например, сотрудники организации должны быть вовлечены в интерпретацию результатов алгоритмов при принятии решений; необходимо сформировать план действий на случай сбоев в процессах принятия решений на основе алгоритмов).

Монетарное управление Сингапура (MAS) разработало [добровольные принципы](#), нацеленные на обеспечение добросовестности, соблюдения этических норм, ответственности и прозрачности при использовании искусственного интеллекта и осуществлении анализа данных (AIDA) в финансовом секторе. В части обеспечения добросовестности выделены необходимость защиты от систематического ущемления интересов физических лиц / групп физических лиц путем принятия решений с помощью технологий искусственного интеллекта, регулярная проверка моделей на корректность работы и исключение систематических ошибок. Модели должны функционировать в соответствии с установленными целями. Под этикой подразумевается, что применение данных технологий должно осуществляться в соответствии с установленными в компании этическими нормами, ценностями и кодексами поведения. Решения, принимаемые на базе AIDA, должны как минимум соответствовать тем же этическим стандартам, что и решения, принимаемые людьми. Принципы, связанные с ответственно-

стью, рассматривают необходимость несения компанией ответственности за применение как разработанных ей, так и полученных из внешних источников моделей. Под прозрачностью понимается необходимость по требованию субъектов данных предоставлять им четкие разъяснения в отношении затрагивающих их интересы возможных последствий решений, принимаемых на базе искусственного интеллекта, а также в отношении того, какие именно данные о конкретных субъектах используются при принятии решений.

Управление комиссионера по персональным данным Гонконга (Office of the Privacy Commissioner for Personal Data) разработало [Руководство по этике](#) (Ethical Accountability Framework, EAF), которое в целом предполагало развитие культуры этичного управления данными и снижение рисков, связанных с использованием персональных данных, в том числе в рамках систем, основанных на использовании больших данных. В 2019 году **Монетарное управление Гонконга** (НКМА) опубликовало [циркуляр](#), в котором призвало финансовые институты беспрекословно следовать принципам EAF, а также [циркуляр](#), направленный на усиление защиты прав потребителей финансовых услуг при использовании технологий больших данных и искусственного интеллекта. Циркуляр содержит ряд руководящих принципов в следующих областях: управления и подотчетности систем технологий больших данных и искусственного интеллекта; справедливого применения систем технологий больших данных и искусственного интеллекта; прозрачности и раскрытия информации; защиты информации.

[Доклад с перечнем аналогичных принципов](#) также был опубликован **Центральным банком Нидерландов** в 2019 году (выделены шесть принципов использования искусственного интеллекта в финансовых услугах: надежность, ответственность, справедливость, этичность, компетентность и прозрачность – SAFEST-принципы).

С 1 ноября 2021 года в **Китае** вступил в силу [закон о защите персональных данных](#): интернет-сервисы и приложения теперь обязаны предоставлять пользователям варианты обработки информации, не требующие их персональных данных, или давать возможность отказаться от предоставления такой информации. Новый закон также обязывает интернет-ресурсы получать согласие пользователей на обработку биометрических данных, сведений о состоянии здоровья, финансовых счетах и геолокации. Данный закон будет оказывать большое влияние на бигтехи (Alibaba, Tencent), которые аккумулируют [значительные объемы информации](#).

Отдельные регуляторы, помимо проработки вопросов, связанных с обеспечением недискриминации клиентов, этичного и ответственного использования технологий обработки больших данных, рассматривают также **вопросы получения информации для регуляторных целей**. Например, в июне 2020 года **Народный банк Китая** объявил о намерении создать [единую информационную базу для финансового сектора](#) на основе технологий обработки больших данных. Данная база данных должна стать единым центром по управлению сбором финансовой информации, ее хранению и анализу. Кроме этого, в феврале 2021 года китайский регулятор опубликовал доклад, который обосновывал необходимость введения **требований к большим интернет-платформам по предоставлению полной кредитной информации регулятору**. Данные о миллионах пользователей интернет-платформ используются большими технологическими компаниями (Ant Group (Alibaba), WeBank (Tencent) и так далее) для предоставления кредитов (при этом сами кредитные средства предоставляются банками, а интернет-платформы в данных операциях фактически являются лишь посредниками).

В июне 2021 года **Комиссия по финансовым услугам Республики Корея** (Financial Services Commission, FSC) обнародовала [планы по развитию аналитики больших данных](#) в финансовом секторе. Указанные меры включают планы по увеличению количества компаний, которые будут специализироваться на **конвергенции данных**; частичному снятию ограничений на конвергенцию данных и их последующему использованию одной и той же компанией при отсутствии конфликтов интересов – согласно текущему законодательству, обработанные данные разрешено использовать только третьим лицам; упразднению ограничения на использование частичных выборок данных; расширению доступа к конвергенции и использованию данных

для МСП, которые не располагают собственными выборками данных, но изъявляют желание использовать данные, предоставленные другими компаниями (это позволит снизить административное бремя для компаний, владеющих исходными данными).

В 2019 году Организация экономического сотрудничества и развития (ОЭСР) опубликовала [рекомендации](#), которые по сути являются первым межправительственным стандартом¹ по искусственному интеллекту. Документ включает пять принципов использования искусственного интеллекта: содействие инклюзивному росту, повышению благосостояния и устойчивому развитию; гарантия справедливого общества; прозрачность и интерпретируемость; надежность и безопасность. Кроме того, необходимо обеспечить адекватную работу искусственного интеллекта в соответствии с вышеизложенными принципами. Правительствам, в свою очередь, рекомендуется инвестировать в НИОКР, связанные с искусственным интеллектом, стимулировать развитие цифровых экосистем для искусственного интеллекта; создать условия, благоприятные для внедрения данных технологий; стимулировать обучение и повышение компетенций работников; работать на международной арене и содействовать международной кооперации разных секторов экономики по вопросам обмена информацией и знаниями, разработки стандартов, внедрения принципов.

5.2. Возможные шаги в области регулирования использования технологий больших данных в России в целях снижения рисков финансовой стабильности

В российском регулировании вопросы использования больших данных затронуты на данный момент лишь в некоторых аспектах.

Так, с целью защиты персональных данных принят Федеральный закон от 27.07.2006 № 152-ФЗ «О персональных данных». Задачей указанного Федерального закона является обеспечение защиты прав и свобод человека и гражданина при обработке его персональных данных, в том числе защиты прав на неприкосновенность частной жизни, личную и семейную тайну.

Положение Банка России от 08.04.2020 № 716-П «О требованиях к системе управления операционным риском в кредитной организации и банковской группе» (далее – Положение № 716-П), Положение Банка России от 06.08.2015 № 483-П «О порядке расчета величины кредитного риска на основе внутренних рейтингов» (далее – Положение № 483-П) и иные нормативные документы Банка России устанавливают требования к управлению банками модельным риском в рамках системы управления операционным риском, во ВПОДК и при применении ПВР в целях оценки достаточности капитала.

Кроме того, в Положении № 716-П указаны требования Банка России к надежности и обеспечению непрерывности функционирования информационных систем, обеспечению качества данных в информационных системах в рамках системы управления операционным риском, требования к обеспечению информационной безопасности, а также требования Банка России, связанные с взаимодействием кредитной организации со сторонними поставщиками услуг с целью соблюдения требований к информационной безопасности и передачей на аутсорсинг элементов информационных систем и данных кредитной организации.

Контроль (мониторинг) рисков, связанных с использованием больших данных, и их оценка должны быть реализованы в рамках системы управления рисками кредитной организации при условии эффективного функционирования системы внутреннего контроля кредитной организации.

С целью адресовать представленные в настоящем докладе дополнительные риски использования больших данных, с учетом возрастающей роли больших данных в банковском бизнесе [Банк России предлагает обсудить следующие меры](#).

¹ К соглашению присоединились 42 страны.

Совместно с банками и иными финансовыми организациями разработать *свод принципов надлежащего использования больших данных*, который мог бы быть в дальнейшем реализован в виде стандартов саморегулируемых организаций, профессиональных ассоциаций участников финансового рынка. В той части, где предметная область указанных принципов пересекается с действующим регулированием, указанные принципы могут служить в качестве развития действующих норм регулирования. На основе опыта использования данного свода принципов и полученных выводов об их актуальности и эффективности в дальнейшем могут быть разработаны рекомендации Банка России в области надлежащего использования больших данных.

Данный свод принципов может также рассматриваться в качестве развития «Национального кодекса этики в сфере искусственного интеллекта»² применительно к финансовым организациям.

В частности, предлагается включить в указанный **свод принципов** следующие пункты:

1. В финансовой организации должна быть разработана стратегия управления данными (иной верхнеуровневый документ), которая бы, в частности, регламентировала вопросы управления большими данными, в том числе порядок их сбора и хранения, обеспечения полноты и качества данных, состав ответственных лиц (подразделений), ответственность руководства организации, порядок осуществления проверок со стороны внутреннего аудита, с учетом в том числе требований Положения № 716-П и приложения 3 к Положению № 483-П.
2. С целью ограничения модельных рисков и рисков качества больших данных в организации должна быть разработана система обеспечения полноты и качества данных, так, чтобы не могло возникнуть ситуации, когда модели, разработанные на основе больших данных, из-за неполноты данных или недостаточного качества самих моделей дают смещенные, некорректные результаты или результаты, ведущие к дискриминации. Должна быть разработана система регулярной актуализации больших данных в соответствии с характером данных. Модели должны проходить регулярную калибровку и независимую валидацию. Сами модели должны быть прозрачными, интерпретируемыми и пригодными для независимой проверки со стороны внешнего аудита и регулятора. Система внутреннего контроля организации должна осуществлять проверку системы управления данными и модельным риском.
3. С целью ограничения рисков, связанных с этическим использованием больших данных, должны быть разработаны специальные этические принципы, которые могут быть закреплены на уровне стандарта СРО.

В частности, указанные этические принципы должны покрывать вопросы использования больших данных в маркетинге (CRM) и ценообразовании, в особенности данных, полученных из внешних источников, и результатов моделей, разработанных на основе этих данных. В том числе должны быть регламентированы вопросы обеспечения сохранности, конфиденциальности и этического использования в отношении внешних больших данных, таких как данные социальных сетей, поисковых запросов в Интернете, операторов сотовой связи, платежные данные, данные геолокации и прочие (которые не являются обезличенными). Под этическим использованием понимается, в частности, **отсутствие какой-либо дискриминации, как ценовой, так и иной, включая гендерную, расовую, национальную и прочую, отсутствие ограничений свободы выбора финансовых услуг**, соблюдение прав и свобод человека (в том числе отсутствие злоупотреблений частной информацией пользователей, размещенной в Интернете). Кроме того, указанные принципы должны регламентировать порядок использования больших данных в директ-маркетинге (таком, как «холодные» звонки, рассылка так называемых «предодобренных предложений» и прочее).

² «Национальный кодекс этики в сфере искусственного интеллекта», подписанный 26.10.2021 в Москве ведущими технологическими компаниями в рамках I международного форума «Этика искусственного интеллекта», был разработан в рамках реализации Национальной стратегии развития искусственного интеллекта до 2030 года, утвержденной Указом Президента Российской Федерации от 10 октября 2019 года № 490. Кодекс разработан при участии «Альянса в сфере искусственного интеллекта». Положения Кодекса могут быть расширены и конкретизированы для отдельных секторов экономики с учетом их особенностей (пункт 2 Кодекса «Допустимость отраслевых кодексов»).

4. С целью управления риском зависимости от сторонних поставщиков больших данных³ или услуг по их обработке во внутренних документах финансовой организации должны быть предусмотрены соответствующие подходы, в частности по мониторингу доли каждого отдельного поставщика больших данных (а также моделей на их основе, услуг по хранению и обработке), стресс-тестированию операционного риска в случае сбоев в работе поставщиков, порядку восстановления деятельности в случае сбоев. Должен быть учтен также страновой аспект риска путем диверсификации иностранных поставщиков, использования услуг национальных поставщиков при наличии возможности. Кроме того, организация должна контролировать качество данных и моделей, получаемых от внешних поставщиков, предусмотреть ответственность поставщиков за качество данных. В любом случае организация должна нести конечную ответственность за применение как разработанных ею, так и полученных от внешних поставщиков моделей.

В дальнейшем предполагается проработка со стороны Банка России вопроса об установлении требований к управлению риском аутсорсинга в рамках системы управления операционным риском, а также внедрения требований к предварительному уведомлению регулирующего органа об аутсорсинге значимых функций для целей идентификации рисков сторонних поставщиков услуг, включая риски концентрации, и предупреждения возникновения в связи с этим новых системных рисков.

В целом полагаем целесообразным определять подходы к управлению рисками, связанными с использованием больших данных (например, модельным риском, риском информационных систем, риском информационной безопасности, риском аутсорсинга), с учетом требований к управлению операционным риском.

³ Не относится к БКИ как к поставщику данных.

ВОПРОСЫ ДЛЯ КОНСУЛЬТАЦИЙ

1. Видите ли вы необходимость в разработке стратегии работы с большими данными на уровне организации? Какие вопросы такая стратегия должна включать?
2. Какие подходы используются (планируются к использованию) в вашей организации с целью управления модельными рисками и рисками качества данных в отношении больших данных?
3. Какие подходы используются (планируются к использованию) в вашей организации с целью управления рисками дискриминации (в том числе неценовой) при принятии решений с использованием больших данных?
4. Каким образом в вашей организации осуществляется управление риском зависимости от сторонних поставщиков услуг в части больших данных и моделей, разрабатываемых на их основе? Видит ли ваша организация повышенный уровень концентрации на каких-либо поставщиках? Какова максимальная доля отдельного поставщика больших данных и услуг на их основе в ключевых направлениях деятельности вашей организации?
5. Поддерживаете ли вы создание упомянутого свода принципов надлежащего использования больших данных финансовыми организациями? Какие ключевые пункты и принципы ваша организация считает целесообразным включить в указанный свод?
6. Какие иные меры регулирования в области использования больших данных ваша организация считает наиболее актуальными и целесообразными для разработки и имплементации?
7. Какую информацию ваша организация хотела бы получать в обезличенной форме для совершенствования использования больших данных?
8. Возможно ли, по мнению вашей организации, закрепление принципов работы с большими данными, в том числе этических принципов, на уровне стандартов СРО и профессиональных ассоциаций участников финансового рынка?