Bank of Russia

USING BIG DATA IN THE FINANCIAL SECTOR AND RISKS TO FINANCIAL STABILITY

Consultation paper

Moscow
2021

CONTENTS

This paper was prepared by the Financial Stability Department.
Please send answers to the questions raised in the report as well as your comments or suggestions no later than 15 February 2022 to yudinats@cbr.ru; vorozhtsovpa@cbr.ru; mmm1@cbr.ru; burkovayua@cbr.ru.

Cover photo: Shutterstock/FOTODOM
12 Neglinnaya St, Moscow 107016
Bank of Russia website: www.cbr.ru

INTRODUCTION

In recent years, the use of big data technologies in the financial sector has become more common. These technologies have the potential to improve the quality of services provided and help financial institutions to reduce their costs and improve performance.

Big data is used by practically every major financial institution, including consulting and technology companies providing services in the financial sector. The Bank of Russia survey conducted in summer 2021 demonstrated that this is also the case for the Russian financial market. With the help of such data, financial institutions can, first of all, more accurately assess the preferences of financial services consumers, make products more personalised, make more accurate assessments of risks and improve compliance and anti-money laundering and combating the financing of terrorism (AML/CFT) procedures.

However, the use of big data may pose a number of risks, including systemic risks for the financial market. Systemic risks may be associated with model risks that can lead to a mass incorrect assessment of the financial situation of borrowers, risks of consumer discrimination (price and non-price), risks of distortion of competition, risks of critical concentration of data providers (including foreign ones) and risks of major leaks of personal data. In this respect, in recent years in different countries, questions of regulating the use of big data by the financial sector have been raised.

This report describes the trends in the use of big data globally and in Russia and analyses the benefits and risks of its use as well as potential approaches to limiting the risks of using big data in Russia, including on the basis of foreign regulatory practices. However, this report does not cover the matters of using and protecting personal data in detail.

1. TERMS AND DEFINITIONS

The term *'big data'* does not currently have a single generally accepted definition. This term is used to refer to large and/or complex data arrays and related technologies for their storage and processing. These data arrays can have either a structured form (external and internal databases) or an unstructured one (social networks, press etc).

The most widely used definition of big data is as follows.

Big data is high – volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision-making and process automation[1].

Thus, the definition of big data implies the presence of three essential criteria: volume, velocity and variety, or the 3Vs[2], where:

- volume means a vast amount of data, often from different sources;
- velocity means high speed of data acquisition, collection and analysis;
- variety refers to the various types of data used, both structured and unstructured.

According to the approach of the UNECE[3], different types of big data can be classified as follows:

- information from social networks, blogs and internet posts;
- data about internet activities (including user search queries and data about sites visited);
- information from traditional business processes (information about transactions, purchases, orders, payments, customer registration, banking operations etc);
- data of public organisations (administrative data, including customs data, tax data etc, and medical data);
- data from mobile and other devices (geolocation data, traffic data, data from smart home systems, video surveillance cameras, data from sensors, trackers and so on).

Since the processing of big data generally takes place using artificial intelligence technologies, the following definitions are used in this respect.

Artificial Intelligence (AI) is a set of technology solutions making it possible to simulate human cognitive functions (including self-learning and searching for solutions without a predetermined algorithm) and to obtain results when performing specific tasks comparable at least to the results of human intellectual activity. The set of technology solutions includes information and communication infrastructure, software (including software using ML methods) and processes and services for data processing and searching for solutions[4].

Machine learning (ML) is a subcategory of AI; the ability of computer systems to extract knowledge from data and use it for subsequent problem-solving[5].

AI technologies are technologies based on AI, including computer vision, natural language processing, speech recognition and synthesis, intelligent decision-making support and advanced AI techniques[6].

---

[1] *Gartner IT Glossary*.

[2] *Sometimes additional criteria are noted, such as veracity, value and volatility (a characteristic indicating how long the data is useful/relevant). In this case, the concept can be called the 4 Vs, 5 Vs and so on.*

[3] *United Nations Economic Commission for Europe.*

[4] *Decree of the President of the Russian Federation No. 490, dated 10 October 2019, 'On the Development of Artificial Intelligence in the Russian Federation' (together with the 'National Strategy for the Development of Artificial Intelligence for the Period up to 2030').*

[5] *The definition is based on various sources, including materials from the report Artificial intelligence and machine learning in financial services, Financial Stability Board, 2017.*

[6] *Federal Law No. 123FZ, dated 24 April 2020, 'On Conducting an Experiment in Establishing Special Regulation to Create the Necessary Conditions for Developing and Implementing Artificial Intelligence Technology in Moscow, a Constituent Entity of the Russian Federation and a City of Federal Importance, and Amending Articles 6 and 10 of the Federal Law 'On personal data''.*

Big data and AI are closely interrelated. Big data provides a source of information for analysis using AI. ML technologies based on neural network approaches generally provide the fastest and most comprehensive results of big data processing. However, they bear certain risks associated with the transparency and interpretability of the respective models. Big data can be analysed by means other than AI. For a certain range of tasks, traditional statistical (econometric) models can be used effectively; however, it is AI that makes it possible to take full advantage of big data processing.

2. TRENDS IN THE APPLICATION OF BIG DATA BY FINANCIAL INSTITUTIONS OF LEADING COUNTRIES

Over recent years, financial institutions have been increasingly using big data and methods for their processing in their activities. However, there are differences both in the scale of use of big data among different segments of the financial market and in the approaches to their use. This chapter analyses the main trends in the use of big data by foreign financial institutions: the overall scale of big data use, the nature of interaction with third-party providers and trends in the applied use of big data technologies for various purposes.

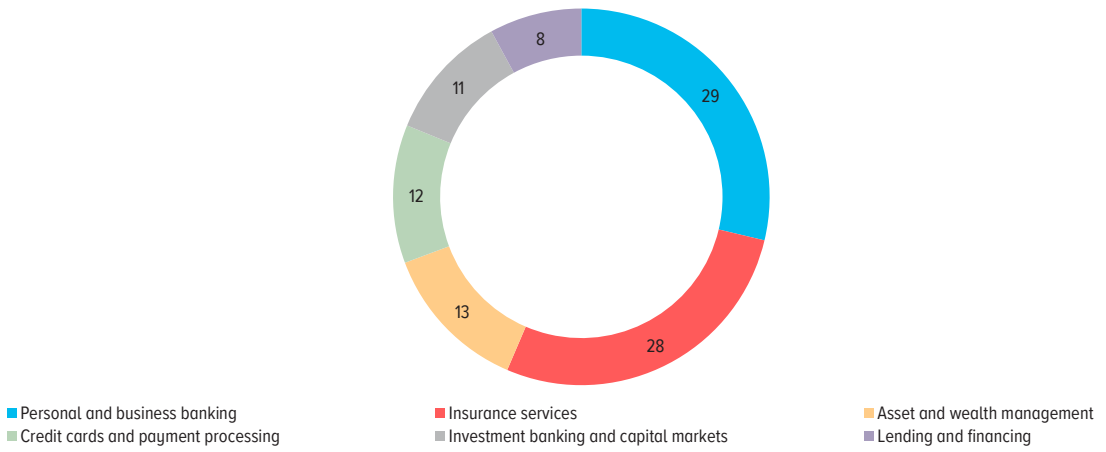
2.1. The main trends in the use of big data

1. **The financial sector has been increasingly using big data and AI technologies for processing it.** This is facilitated by an increase in the volume of information in general: households and companies produce an unprecedented amount of data that is constantly accumulating, social networks make it possible to track the speed of dissemination of information, companies document the process of the production and sale of goods and services in detail, payment transactions and e-commerce also have their own digital footprint. Investments in big data from financial sector companies are growing; according to some estimates, they will amount to $14 billion in 2021 versus $9 billion in 2018.

According to a study by the ***European Banking Authority (EBA)***, EU banks are interested in using big data analytics: there is growth both in the share of organisations already using big data in their activities (from 60 to 64% from 2018 to 2019) and in the share of companies testing, developing and discussing the introduction of new technologies. Only 2% of banking institutions said that they are not implementing big data analytics.

2. **The Covid-19 pandemic has probably accelerated the adoption of big data technology as well as AI/ML for big data processing.** Typically, during a crisis, business cuts back on investments in innovative technologies and prioritises short-term cash needs over long-term technology projects. However, in August 2020, ***the Bank of England*** conducted a study on banks' use of ML, AI and big data amid the Covid-19 pandemic[1] which revealed a completely different trend: the

FINANCIAL SECTOR INVESTMENTS IN BIG DATA BY TYPE OF FINANCIAL SERVICES *Chart 1*
(%)

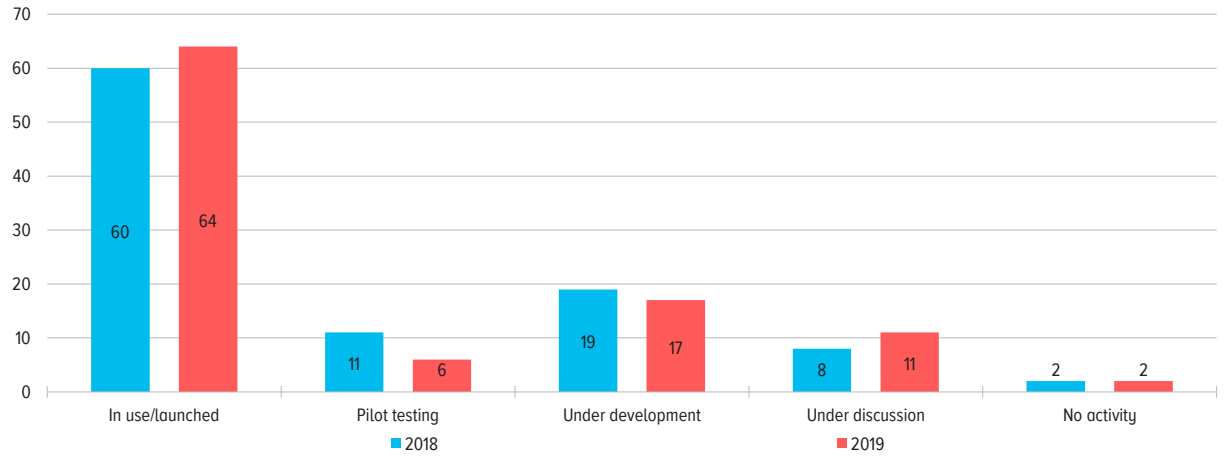


Source: World Bank.

[1] *The impact of Covid on machine learning and data science in UK banking*.

USE OF BIG DATA ANALYTICS ACROSS EU INSTITUTIONS
(% OF INSTITUTIONS SURVEYED)

Chart 2



Source: European Banking Authority (EBA).

total investments of banks in big data technologies increased during the pandemic. According to the study, this situation arose for a number of reasons:

- the need for digitalisation of business has increased overall given the transition of consumers to online channels during the pandemic;
- the desire of financial institutions to improve performance through process automation and in the context of low interest rates;
- the financial services sector was not, in general, affected as badly as other industries during the pandemic;
- an increase in the amount of data amid more extensive use of computers and smartphones for commerce, teleworking and communication by the population during the pandemic.

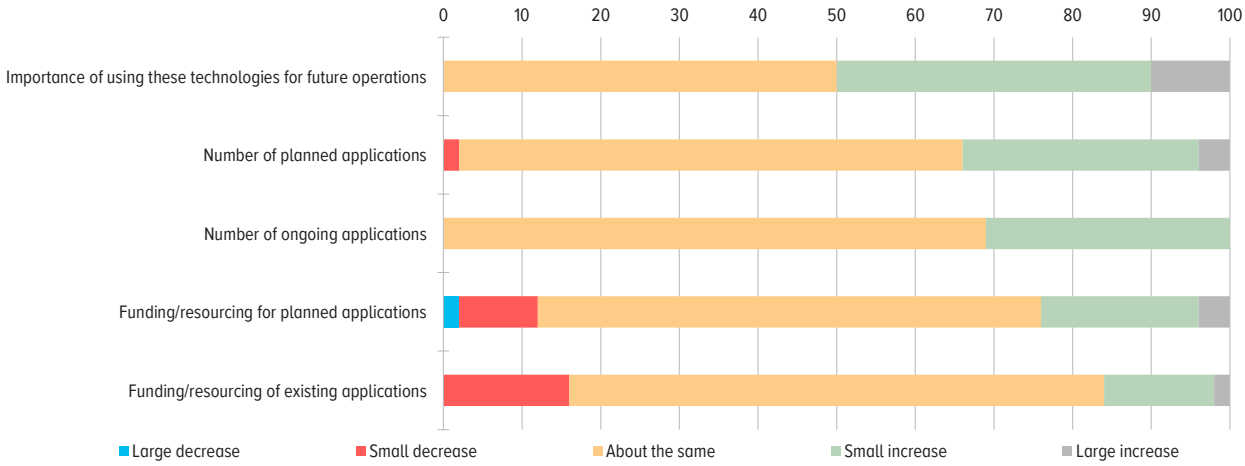
As banks adjust costs and revenues amid the Covid-19 pandemic, they are looking to leverage ML and other techniques for processing large volumes of data to enhance performance and improve digital channels for reaching customers. The pandemic has spurred interest in ML and other methods of processing significant volumes of data in the UK banking sector. Half of the banks surveyed reported an increase in the importance of big data technologies for their future operations, and none of the banks reported a decrease. In terms of investments, an equal number of respondents (16% each) wrote of a decrease or an increase in the amount of resources allocated to existing big data processing programmes; regarding the development of new programs, 23% of banks reported an increase in allocated resources, and only 11.5% of banks reported a decrease.

An example of the use of big data technologies amid the Covid-19 pandemic is the creation by the Indonesian Bank Mandiri together with the American developer Cloudera[2] of the Enterprise Information & Decision Platform, a special big data platform which makes it possible to monitor the liquidity and daily operations of its branch offices in real time with a visual map of Covid-19 affected areas, on the basis of which a decision is made on which branch should continue to operate and which should terminate its operations. The platform also makes it possible to submit relevant information to customers and the regulator in a timely manner, to analyse factors that can facilitate the process of restructuring customer debts and to monitor the health of bank employees on a daily basis, which minimises the risk of infection.

3. **In general, when handling big data, small financial institutions tend to use third-party developments, while large financial institutions have sufficient resources to develop their own solutions.** According to the EBA study, many European financial institutions use third-party developments (e.g., modelling tools); however, responsibility for the quality of models created on the basis of developments of third-party suppliers and their results lies with the financial institutions themselves.

[2] *Collaboration with Cloudera started back in 2017.*

IMPACT OF THE COVID-19 PANDEMIC ON THE CURRENT USE AND PLANS FOR THE INTRODUCTION OF MACHINE LEARNING AND OTHER METHODS OF PROCESSING SIGNIFICANT AMOUNTS OF DATA *Chart 3*
(% OF BANKS SURVEYED)



Source: The Bank of England.

At the same time, a study by the Bank of England and the UK Financial Conduct Authority (FCA) found that British financial companies mainly use their own programmes for processing data through ML; 76% of respondents said that programmes for using ML had been developed and were being used within their companies. At the same time, when creating their own data processing programmes, financial companies often use elements developed by third-party vendors (models, software and ML libraries), which are subsequently adapted for internal use. Most often (36% of cases), non-bank financial institutions use the developments of third-party vendors. Perhaps this is due to the small size of such companies and the lack of resources for their own developments.

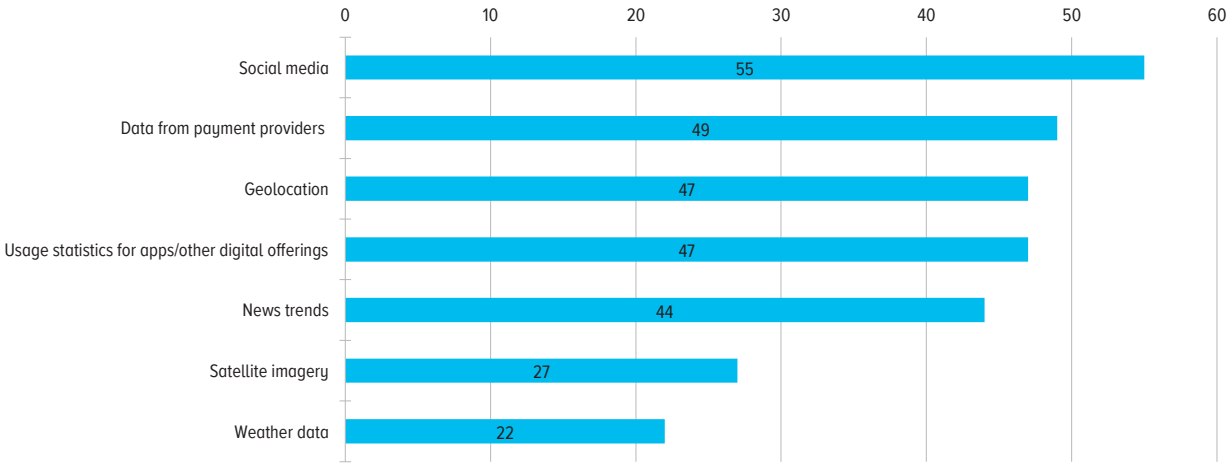
Moreover, financial institutions are increasingly using cloud storage services if a shortage of in-house server capacity is observed. According to a study by the Bank of England and the FCA, 22% of applications used by financial institutions for processing ML data run on cloud servers. It is emphasised that applications developed by a financial institution itself generally run on internal servers. Cloud services are more often used by non-bank financial institutions (as mentioned above, they also more often use the services of third-party providers when it comes to developing applications for data processing).

An example of the use of services of third-party technology providers for processing big data is the agreement between Deutsche Bank and Google Cloud for the development of high-tech solutions based on the use of ML and AI, which is expected to help the bank improve risk analysis, cash flow forecasting and decision-making on security solutions. Moreover, under the agreement, Google Cloud will provide the German bank with cloud storage services for big data. One of the largest Turkish banks, Akbank, uses the American cloud data processing platform Datameer. An IBM platform enables the Brazilian bank Banco do Brasil to process big data in its risk management process.

4. **In addition to traditional sources of information, financial institutions are using alternative sources of data, from social media to satellite data.** This approach provides access to previously unaccounted-for data that can be used for various purposes, including trading in securities and data-driven investments (for example, when making automated investment decisions in the stock market or performing credit scoring, to improve the quality of automated algorithms using ML).

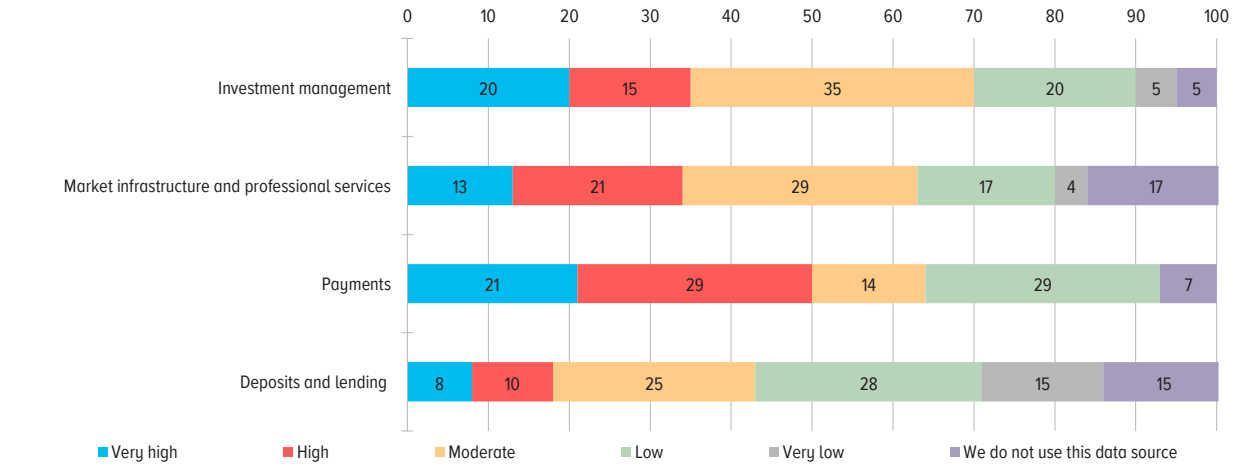
According to an EBA study, for European banks, proprietary data (information on customer transactions, data on the use of banking products, data on debt service) is the main source for analysis. External data (social media data, public data on real estate prices, information in mass media etc) raises concerns among some banks regarding reliability and accuracy since analytics using low-quality data can lead to poor decisions. According to a study by the Bank of England and

COMMON SOURCES OF EXTERNAL DATA *Chart 4*
(% OF SURVEYED FINANCIAL INSTITUTIONS USING THE PROPOSED EXTERNAL DATA SOURCES)



Source: World Economic Forum.

USAGE LEVELS OF EXTERNAL CUSTOMER DATA IN AI APPLICATIONS *Chart 5*
(% OF FINANCIAL INSTITUTIONS SURVEYED)



Source: World Economic Forum.

the FCA, the surveyed financial companies use external data when applying ML technologies in 40% of cases; data from various sources (consumer information for assessing credit risk, information about cars for insurance and so on), which can be analysed in combination with proprietary data to improve the quality of forecasts, is used.

In 2019, the University of Cambridge and the World Economic Forum conducted a study on the implementation of AI technologies in the financial sector, which covered 151 companies across 33 countries (54% of them are fintech companies, and 46% are traditional financial institutions). The study, in particular, analyses the main sources of external data: the most frequently used sources of external data are social media, data from payment providers and geolocation data (they are used by 55, 49 and 47% of financial institutions, respectively). The least popular sources are satellite imagery and weather data, which tend to be more expensive and may require specialised skills for their processing and interpretation.

External data is most popular among payment providers (50% of companies in this sector answered that they factor in external data to a large extent) and investment managers (35% of investment managers use external data on a large scale). External data is least popular among deposit and lending sector (only 18% of respondents in this segment shared that they extensively use external data).

An important source of big consumer data is data from large high-tech companies (bigtech). This data is then used by associated financial high-tech companies (fintech). Besides increasing the customisation of financial services and improving the options for assessing borrowers, this raises the issue of competition and necessitates government regulation (see subsection 3.2).

2.2. Areas for the use of big data

1. Customer credit scoring

Customer credit scoring is a traditional area for using big data. Banks can improve the quality of credit scoring of new customers by using sophisticated analytical models with significant amounts of data held by financial institutions (sometimes in combination with external data). Application programming interfaces (APIs) can be used to access payment information to offer new services to customers, such as instant lending or pre-approved loans to new customers.

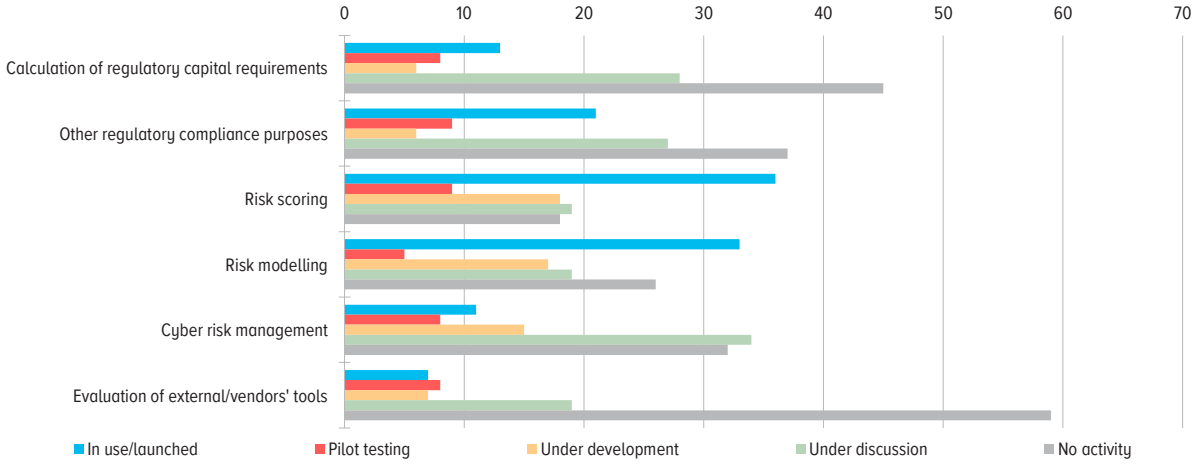
A study by the Bank of England and the FCA found that 57% of financial institutions surveyed use ML technologies in their risk management models.

As mentioned above, big data, including data held by bigtechs, is increasingly being used for retail customer credit scoring. In China, companies using data on online purchases to score the creditworthiness of borrowers (including small and medium-sized enterprises (SMEs)) using big data have begun to emerge: Alibaba launched MYbank, an online bank that uses data from Alibaba's customer base. In 2015, the Chinese company Tencent launched WeBank, an internet bank that uses data from online purchases, social media activity and online gaming for credit scoring.

In 2018, the five largest Brazilian banks [3] created Quod, an alternative to a credit bureau which calculates borrower credit risk taking into account the presence of overdue payments and compliance with the repayment schedules for existing loans. The project uses big data analytics. The technical side of the project was the responsibility of the American company LexisNexis Risk Solutions.

Chilean fintech company Destacame launched a credit scoring platform using alternative data (utility bills and other socio-economic information) to assess the creditworthiness of borrowers in Chile and Mexico who are generally denied financial services by banks. The company cooperates with traditional financial institutions, providing them with an alternative credit scoring of borrowers and serving as a kind of intermediary between the financial institution and the borrowers.

USE OF BIG DATA BY EUROPEAN BANKS FOR RISK MANAGEMENT PURPOSES *Chart 6*
(% OF FINANCIAL INSTITUTIONS SURVEYED)



Source: European Banking Authority (EBA).

[3] *Banco do Brasil, Bradesco, Itaú, Santander and Caixa Econômica Federal.*

It should be noted that sometimes big data platforms for improving the availability of financial services are created not by the private sector but by regulators. In 2020, the Hong Kong Monetary Authority announced its intention to establish a new financial market infrastructure, the Commercial Data Interchange (CDI), which aims to improve the accessibility of financial services in Hong Kong for SMEs by ensuring the safe and efficient exchange of information between banks and sources of commercial data about SME transactions. CDI will provide banks with verified data on the behaviour of potential borrowers (including payment data etc) that can be used by banks for credit scoring. Data will only be provided to banks with the consent of the organisations.

2. Marketing and customer interaction

With the help of big data, companies can track various aspects of customer behaviour and receive information about their preferences and then, based on the obtained data, send relevant offers of certain bank products which are more likely to be of interest to customers since they will correspond to their individual needs. Financial institutions can use big data from various customer interactions (including handling complaints) to more accurately identify and fully meet customer needs, including with the help of chatbots. Moreover, analysing sales, products and the market in general using big data can help develop new financial products that meet customer needs and ensure a competitive advantage in the market.

In its study, the EBA notes that nearly 95% of banks surveyed are in the process of exploring the use of big data to interact with customers or are already using big data analytics (about 50% of companies).

In the French banking system, digital transformation affects mainly the major players in the industry. The latter serve as an incubator for startups operating based on a neobank model[4]. For example, within one of the largest French conglomerates, Société Générale, the neobank Prismea was created, which offers digital banking services to SMEs and uses big data and AI to provide customised consulting services.

UK-based Metro Bank[5] used AI and predictive analytics to help its customers manage their finances. Working in partnership with Personetics, the bank launched an in-app Insights service that monitors customers' data and transaction patterns in real time. The app then provides personalised payment tips and alerts for unusual charges.

The Canadian financial group Canadian Western Bank, in partnership with fintech company Temenos Infinity, provides personalised solutions for SMEs to efficiently manage their cash flows and financial decisions.

In 2018, HSBC launched the Business Internet Banking platform in Hong Kong, which analyses data on all transactions in the foreign exchange market made by the bank's customers and provides tips and advice. For example, if a customer performs several operations of the same type for a certain time, it may offer to automate the process.

3. Asset management

Big data analytics tools such as AI and ML are increasingly being used in asset management. The French banks BNP Paribas and Société Générale are working to implement AI for big data analytics to optimise the performance of their portfolios. Société Générale is testing ML models to build portfolios of stocks with a target return-risk ratio. The model uses big data combining 80 financial indicators (stock price, issuer's EBITDA, profitability ratios etc) and non-financial ones (ESG characteristics, expert comments, media information).

---

[4] *Fintech companies offering standard fully digital banking services and having no physical branches.*

[5] *AI leaders in financial services: Common traits of frontrunners in the artificial intelligence race.*

In turn, BNP Paribas Asset Management has implemented ML technologies for scoring issuers based on analysis of the time series of their operating and financial indicators and stock quotes and also uses natural language processing (NPL) and sentiment analysis technologies to evaluate media statements concerning the activities of companies.

The Anglo-German company Arabesque Asset Management has developed a big data platform for assessing ESG factors for more than 7,000 of the world's largest listed companies. More than 200 ESG indicators from over 30,000 sources across 170 countries are analysed on the basis of self-learning algorithms. Arabesque not only uses this development itself when forming an investment portfolio but also provides access to the platform for other companies.

4. Insurance

Along with banks, the insurance sector is a leader among financial institutions in implementing big data. Many insurance companies use AI and big data to automate claims processing, detect fraud and ensure regulatory compliance.

Moreover, in vehicle insurance, big data technology is used in conjunction with the Internet of Things, where the main source of data for analytics is telematics devices (black boxes) making it possible to determine how a vehicle is driven and, consequently, provide a more accurate risk assessment. Data collected in real time using telematics devices allows insurance companies to clarify the customer risk profile and, as a result, optimise the cost of insurance services.

Insurance companies view big data as the backbone of their future business models and pricing policies. In France, a number of insurance groups have begun implementing omnichannel strategies[6] and leveraging big data and AI technology to better analyse sales, including forecasting sales of insurance products.

In February 2020, the Australian Prudential Regulatory Authority (APRA) conducted a survey of 36 regulated insurers on new technologies used in insurance (InsurTech), including big data, AI, software as a service, blockchain and the Internet of things (including telematics). According to the survey, big data is InsurTech's most popular category, used primarily for pricing and underwriting.

5. Prevention of materialisation of operational risks, including cyber risks and risks of AML/CFT violation

Financial institutions can use big data to reduce costs and improve the efficiency of their operational risk management procedures and compliance systems. For example, the analysis of large amounts of data can help simplify the process of fraud detection. An important area of application of big data analysis is ensuring cyber security (analysis of cyber threats, identification of suspicious activities that can disrupt the operation of the bank's internal systems).

In 2018, HSBC launched its Global Social Network Analytics platform to detect and prevent financial crime using big data. The platform analyses, including in real time, both HSBC's own data and external data from various sources (for example, information on the ownership structure of companies or assets, which makes it possible to identify links between counterparties and transactions). The platform analyses transactions taking into account 50 different scenarios, trying to identify any signs of money laundering, for example, hidden connections between the participants in a transaction or suspicious payment dynamics. The technology solution for big data analysis and external data for the HSBC platform is provided by the British company Quantexa. However, the data for Global Social Network Analytics is stored on HSBC's internal servers.

[6] *Omnichannel marketing based on big data involves collecting and unifying large amounts of data to create detailed customer profiles, segment them and use business intelligence to optimise the marketing strategy.*

6. Optimisation of reporting and other processes

Big data is increasingly being used to optimise and improve operational efficiency in daily operations of financial institutions, in particular, for conducting transactions, interacting with regulators and improving employee performance. Some financial institutions, such as Citibank, use big data to automate their reporting to regulators. German Commerzbank uses AI to automatically check trade finance transactions for compliance with legislation before they are actually performed as well as to prepare profit and loss statements. Another German bank, Deutsche Bank, has implemented various technologies for processing big data on trade operations and risk.

Moreover, big data can be used to analyse possibilities for improving employee performance. Hovered, the EBA notes that currently there is limited use in this area.

At the beginning of the last decade, Singapore-based DBS bank began using big data analytics to improve its personnel management efficiency and reduce employee turnover. An analytical department was created within the HR department to assess trends and identify risk factors for personnel turnover. Over eight years, the personnel turnover rate dropped from 27% to 18% (below the market average).

Given the above trends, it can be concluded that in the financial sector interest in technologies related to the use of big data will continue to grow in the future against the backdrop of an increase in the volume of data produced and due to the wide possibilities for its use in various aspects of financial institutions' activities.

3. BENEFITS AND RISKS OF USING BIG DATA

Big data can be used in various areas of financial institutions' activities, making it possible to increase performance, reduce costs and increase profits. Due to large volumes of information processed, the speed of data processing and the increasing level of automation, big data technologies provide financial institutions with significant advantages in risk management and help optimise operational activities and business processes and improve the operation of the financial market. Moreover, big data enables financial institutions to offer more personalised services to consumers and expand their customer base. Nevertheless, the use of big data technologies may pose a number of risks both for financial institutions and for consumers of financial services and the financial system as a whole: methodological risks, risks of improper protection of personal data, risks associated with competition distortion and discrimination, risks of third-party service providers and grey areas in regulation.

3.1. Benefits of using big data

Based on the areas of use of big data analysed above, it can be concluded that big data together with technologies for processing it creates a number of important advantages for financial organisations.

a) ***The high speed of processing large volumes of information*** is one of the most important advantages of big data technologies, which helps improve various aspects of financial institutions' activities. Big data technologies make it possible, first, to quickly extract information from various external sources (social media, the press and other sources of unstructured data) and, second, to analyse external or internal data in real time, including by means of ML. The high speed of data processing, generally without time lags, helps improve forecasting and analysis and make financial decisions that previously took a significant amount of time (money transfers, decisions on issuing loans, approving the issuance of insurance policies) much faster, providing consumers with almost instant access to financial products.

   In a fairly technologically advanced stock market that extensively uses data analysis (usually for structured data, for example, stock price dynamics), big data increases the speed of operations, including through even greater automation and wider use of algorithms for making investment decisions, predicting future prices, portfolio management and determining service fees.

b) ***The quality of risk management*** is improved thanks to the ability to analyse a significant volume of data (including unstructured external data that was not previously taken into account). The use of AI technologies helps identify patterns and relationships, for example, in consumer behaviour, that cannot be detected using traditional tools. The use of big data in credit scoring makes it possible to obtain a better risk assessment, including when issuing loans, which contributes to more equitable (in terms of risk) rate setting. In addition, these technologies can help improve the efficiency of risk modelling and management of operational risks, risks of AML/CFT violations and cyber risks by tracking customer behaviour for suspicious activity using AI.

c) ***Improving access to financial services:*** in many developing and emerging markets, households and SMEs have limited access to loans due to the fact that banks often lack data to assess the creditworthiness of potential borrowers. For example, according to some estimates, most people in China still have difficulty accessing loans, while SME loans account for only 20–25% of corporate lending, although the sector's contribution to GDP is about 60%, and it's account for 80% of total employment. The use of big data can help eliminate this problem

as borrower credit scoring may be based not only on information about loan payments but also on information about the compliance with payment schedules for rent, utilities, mobile communications etc.

Big data can help banking financial institutions increase their customer base on account of households and small- and medium-sized businesses that do not have a credit history, but whose creditworthiness can be analysed using alternative data sources (rent and utilities for households and sales for companies)[1], which also increases the accessibility of financial services throughout the financial system.

On the other hand, in the event of the use of incorrect or incomplete big data and/or incorrect models based on them (model risk), the accessibility of financial services for some borrowers may not increase but may even decrease as a result of an incorrect assessment of their creditworthiness, as, for example, was shown in a study of the consumer lending market in the United States[2]. Moreover, the accessibility of financial services may decrease for some categories of borrowers as a result of discrimination on a certain basis using models based on big data (for more details about model risk and discrimination risks see subsection 3.2).

d) ***The possibility of offering personalised services to a larger number of customers:*** whereas previously identifying customer needs required their personal communication with representatives of a financial institution, now the analysis of big data helps provide personalised services to millions of customers. Analysing consumer spending, social media behaviour and geolocation potentially enables financial institutions to better understand customer needs. Moreover, with the help of big data, companies can conduct more targeted marketing and pricing (which is especially relevant for the insurance sector) through more accurate segmentation of potential customers. Using big data for a more personalised approach is beneficial both for customers given the improvement in the quality of services provided and for financial institutions themselves (customers prefer companies with better services).

Processing large amounts of customer data helps not only to better meet customer needs using existing products but also to identify what financial products are lacking in the financial market, which can be used to form new products for customers. The creation of new products based on a preliminary analysis of customer needs using big data will have a positive effect on customers (improving the quality of services provided), the financial institutions themselves (increasing profits) and the development of the financial system as a whole.

Thus, in general, big data technologies contribute to better processing of information in various segments of the financial market, which can potentially reduce the costs and risks of financial institutions and improve the efficiency and stability of the financial market as a whole. In terms of financial stability, the use of big data may reduce risks by improving the quality of risk management and customer relations processes.

3.2. Risks associated with the use of big data

1. When big data is used in risk management and optimisation of operational activities, ***methodological risks arise, including risks associated with data quality***. Big data analysis methodologies are still under development. It is still not clear how best to select, process and aggregate data. Also, there are questions about the analytical tools required to integrate big data analysis results with information from traditional sources.

   It should be noted that the use of external data related to text recognition and analysis of relationships by processing unstructured information from the media, social networks and other sources requires new automated approaches to data quality management to detect distortions of facts or misinformation (fact checking).

[1] *However, financial institutions should consider the possibility of model risk.*

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

To increase the level of completeness and the quality of big external data, it is crucial that the data used come from independent sources.

2. Another related problem is ***model risk*** when building models based on big data: using erroneous input data or assumptions, using the model for a purpose other than that intended upon its development or errors in the development of the model itself. The fact that AI is often associated with the black box problem presents certain challenges: data is put into the model, and AI conducts an analysis and produces a result that cannot be double-checked or interpreted[3]. In this respect, the issue of the qualification of employees working with big data is important. At the operator level, in order to minimise errors and maximise the options provided by big data, specialists must understand the influence of the initial parameters used in the models and their interrelationships and evaluate the potential impact of parameters not included in the model. An incomplete understanding of the functioning of various forms of big data analytics and the quality of data or models may lead to the materialisation of model risk. Top management, in turn, must have sufficient knowledge to understand the modelling results and the applicability of the models to the tasks for which they are used in the financial institution.

   The materialisation of model risk may lead to such unfavourable consequences as systematically incorrect risk assessment when using big data in risk management. Customer solvency may be assessed incorrectly when conducting credit scoring, and such errors may be repeated on a wide scale, which will lead to the accumulation of credit risks in the banking sector.
3. When using big data ***for customers of financial institutions, there are risks associated with personal data protection***. Big data involves collecting various information about customers. On the one hand, it helps improve the quality of financial services through customisation. On the other hand, the risks of inappropriate use of personal data and the need for its protection increase. The most sensitive is individuals' financial and health data (which can be used primarily by insurance companies). Any potential misuse of big data and lack of information security may undermine consumer confidence in the long run. At the same time, people often share sensitive information about themselves, including in social media, without fully understanding the importance of the information they share and the range of those tasks for which personal data is or can be used.
4. Big data makes it possible to more accurately assess the needs of each consumer individually but also allows financial institutions to assess the willingness of each individual customer to pay a certain price for a service, which may potentially lead to ***price discrimination***. That is, in fact, the same service may have a different price for different customers depending on their ability to pay or other circumstances. This situation leads to a decrease in consumer surplus, especially if a customer needs the service promptly (for example, a loan or an insurance policy) and has not analysed the prices in the market in advance or does not have a penchant for such analysis based on the history of the customer's previous purchases over the internet. Based on the analysis of the IP addresses of the devices which the buyers use to access the internet and geolocation data, financial service providers can track whether a buyer lives in a rich or poor area and offer a different price based on this information.
5. One of the consequences of the lack of transparency or interpretability of methods for processing big data using AI includes ***possible manifestations of (non-price) discrimination on racial, national, religious, gender and other grounds***. Given the lack of transparency of these methods (in contrast to traditional statistical models), discrimination may be unintentional on the part of a financial institution. Since different groups of people have different behavioural characteristics,

---

[3] *It should be noted that in this case we are speaking of artificial intelligence models based mainly on the use of neural networks. However, big data can be analysed not only by means of such methods but also by conventional statistical (econometric) ones. Although in this case model risk is also present, the models are much more transparent and interpretable as they generally show all factors influencing the results (there is no black box problem). However, neural network models are considered the most promising method for working with big data in terms of speed and completeness of results.*

including on the internet, in social media, in the process of online shopping and so on, a model with insufficiently high quality can, based on these features, give a result indicating a lower level of borrower creditworthiness than their real creditworthiness as a result of underestimation of the behaviour characteristics of different population groups. Thus, big data models can be inaccurate and biased in respect of different groups of borrowers.

At the same time, non-price discrimination may, in theory, also be deliberate on the part of a financial institution, again because of the lack of transparency of the methods used when working with big data, including for regulators and auditors.

6. ***The question of the impact that the use of big data can have on competition is controversial.*** On the one hand, the use of big data by large financial institutions, which initially have a significant amount of information about their customers and have sufficient resources to implement new technologies, gives such companies a competitive advantage over smaller market players. Large databases make the best use of the elements of ML and AI for scoring borrowers, assessing policyholder risks or offering customised financial services, while the absence of a large volume of data often makes these technologies ineffective to use. On the other hand, the use of big data can stimulate competition through the entry of fintech and bigtech companies, which often use big data processing technologies, into the financial services market; this creates competition with traditional financial institutions overall.
7. Serious risks to financial stability are ***the risks of third-party providers of big data and services for their processing and the emergence of new systemic risks in this respect***. Since not all financial organisations have the resources or competencies to work with big data in-house, the involvement of new, third-party players (providers of external information, developers of models for analysing big data, cloud services) leads to the complication of relationships in financial markets. Financial institutions may not be able to fully assess or manage these risks as these risks are beyond their organisational structure. The scale of new interconnections may increase the complexity of the financial system and create new channels for the diffusion of systemic risks. In particular, as financial technologies, including big data, gradually erase clear functional boundaries between different types of financial institutions, there is a risk of more rapid spread of contagion effects between sectors and regions. The concentration of third-party suppliers in the market may lead to the emergence of new systemically important players. If the share of several large third-party providers in certain segments of the financial market is high, disruptions in the operations of such companies may lead to large-scale disruptions in other parts of the financial system or the economy as a whole.
8. Moreover, there are concerns about ***the risks associated with grey areas in regulation***. Big data service providers or new types of companies, including providers of big data and models based on them, may fall outside of the full regulatory perimeter of financial regulators[4]. As a result, grey areas of financial activity may emerge that are not subject to regulation. This situation will be especially dangerous if companies outside the regulatory perimeter become systemically important market players.
9. Another challenge is the increased ***number of risks for financial institutions that are lagging behind in the implementation of big data technologies***. Big data is used successfully by a number of financial institutions to monitor and prevent the materialisation of operational risks, including cyber risks and risks of AML/CFT violations. It is expected that more and more financial institutions will introduce these technologies in the future. However, there is a risk that fraudsters will target financial institutions that are lagging behind in the implementation of these technologies (often this means small- and medium-sized financial institutions, which may have insufficient resources to implement new technologies).

[4] *Personal data operators in the Russian Federation in any case fall within the competence of Roskomnadzor.*

Thus, the use of big data technologies by financial institutions has significant advantages but also poses significant risks. It should be noted that some of the risks are associated with insufficient experience in the use of big data by financial institutions (for example, a lack of qualified personnel, methodological risks), which may be eliminated as the market develops. To maximise the benefits of big data technologies, financial market regulators are working on guidelines and requirements aimed at promoting a culture of ethical use of personal data and responsible use of AI and data analytics (see Chapter 5 of this report).

4. USE OF BIG DATA BY RUSSIAN FINANCIAL INSTITUTIONS

In summer 2021, the Bank of Russia conducted a survey of Russia's largest financial institutions regarding the use of big data in their activities. Based on the survey results, the Bank of Russia analysed the main trends and associated risks. The survey included 13 banks (including 9 systemically important banks), 2 credit history bureaus, 1 exchange, 1 large insurance company, 1 payment system and 1 system of money payments and transfers.

4.1. Areas for the use of big data by Russian financial institutions

In the opinion of the overwhelming majority of respondents, in the last 3 to 5 years, there has been an **increase in the use of big data** and in overall interest in it in Russia. According to the majority of respondents (58%)**, the pandemic did not have any noticeable impact on the digital transformation of the business model**; however, according to the rest (42%), the pandemic stimulated the use of big data technologies.

Practically all lending institutions and the vast majority of non-bank financial institutions that took part in the survey included the development of the use of big data processing technologies in **their business strategy**.

All **credit institutions** that took part in the survey reported that they are already quite extensively and effectively using big data in their operations and plan to further expand its use in most of their business processes.

The main area of use is credit scoring. Big data is used to develop risk assessment models for both retail and corporate lending. With the help of big data technologies, pre-approved offers for customers are formed, and the processes of collecting overdue loans are being improved.

Big data is used in risk modelling, customer behavioural analysis as well as for the purpose of information security and counteracting fraudulent actions against bank customers (anti-fraud systems for detecting atypical customer behaviour etc) and in automating responses to customer requests (chat bots).

Big data is used to develop the most active customer relationship management (CRM) channels and cross-selling models and to estimate customer tariffs.

Moreover, as part of working with big data, some banks analyse customer transactions to determine (adjust) the credit ratings of corporate borrowers and individual entrepreneurs.

The monitoring of risks of counterparties which are financial institutions includes analysis of the news background for the timely identification of negative signals about the counterparties' financial stability. News analytics are used to identify the relationships between borrowers. Here, in addition to data analysis methods based on ML, methods and tools for graph analysis are becoming more and more relevant, for example, for studying the affiliation of financial market players or analysing cash flows.

Some banks use big data to build PD/LGD models of borrowers and to calculate the amount of economic capital.

Most of the banks surveyed note that 90 to 100% of retail loan applications are processed using big data technologies, while in the asset portfolio as a whole the share of big data use by various banks accounts for 30 to 90%.

Some banks use robots to build portfolios, including robo-advisors for customers. Robots go through all possible strategies and choose the best one.

Furthermore, big data is used to automate the internal processes of banks.

In terms of the degree of implementation of big data technologies, it may be noted that **among various areas of credit institutions' activities the level of advancement in the implementation of these technologies varies**. For example, in risk management, customer relations, underwriting and pricing, the majority of respondents note large-scale implementation of the respective technologies or implementation in a sufficient volume. At the same time, in terms of market analysis, trading, asset management and payments, most banks are only investigating the matter of introducing technologies based on big data (Table 1).

Among the areas of risk management, banks report using big data mainly in the management of credit risks, especially retail credit risks (Table 2). The assessment of other types of risks (market, liquidity and others) as well as the assessment of capital adequacy is conducted using big data by a smaller number of banks. For example, some banks have already implemented it, while others are only considering it; yet others have no plans for its implementation (Table 2).

Insurance companies use big data analysis in underwriting and to determine demand for insurance products (customer analytics) as well as for the analysis of big data on loss and accident rates to form tariff rules. In **insurance companies**, the share of the use of big data in the formation of their retail portfolio reaches 100%, and for the entire portfolio, 60%.

Exchanges use big data for compliance purposes as well as for analysing customer behaviour and for developing analytics products and offering them to customers.

Payment system operators use big data to plan marketing campaigns as well as for market analysis, risk management and processing payments.

Credit history bureaus use big data to create scoring products, conduct market analysis and develop products.

In general, the majority of organisations (58%) report the use of big data in a wide range of business processes, and only a minority of respondents report its implementation in only one or a few business processes (33%) or the launch of the pilot stage of application of the respective technologies (8%).

USE OF BIG DATA TECHNOLOGIES IN VARIOUS BUSINESS AREAS OF FINANCIAL INSTITUTIONS AND THE DEGREE OF THEIR IMPLEMENTATION (ACCORDING TO SURVEY DATA) *Table 1*

	Not used and not yet planned (%)	Possible implementation is being considered (%)	Development phase (%)	Limited implementation (%)*	Implemented in sufficient volume (%)	Large-scale implementation (%)*
Risk management	0	8	0	17	42	33
Interaction with customers (call centres, creation of personalised offers etc)	0	0	0	33	42	25
Market analysis	17	42	8	17	17	0
Underwriting and pricing	0	17	25	8	25	25
Trading	33	17	17	8	17	8
Asset management	17	33	17	8	17	8
Making payments	25	33	17	8	17	0
Other operational processes, including AML/CFT	10	10	0	80	0	0
Other	0	0	67	33	0	0

THE DEGREE OF IMPLEMENTATION OF BIG DATA TECHNOLOGIES IN THE FIELD OF RISK MANAGEMENT BY FINANCIAL INSTITUTIONS (ACCORDING TO SURVEY DATA) *Table 2*

	No plans for use yet (%)	Possible implementation is under discussion (%)	Development phase (%)	Pilot testing (%)	Used (%)
Capital adequacy assessment, including:	17	33	25	0	25
when using IRB	18	36	27	0	18
when using ICAAP	27	18	27	0	27
Asset valuation, including based on IFRS	9	27	27	9	27
Regulatory compliance, namely:					
capital requirements	55	9	9	0	27
Compliance, including AML/CFT requirements	18	18	18	9	36
Other					
Risk assessment, including:	10	0	0	10	80
retail lending	0	9	9	0	82
SME lending	9	9	9	18	55
corporate lending	18	0	9	18	55
assessment of market risk for a trading portfolio	36	18	18	0	27
operational risk assessment	36	18	9	9	27
liquidity risk assessment	36	18	9	9	27
including assessment of interest rate risk for a banking portfolio (ALM)	27	18	9	9	36
other risks	0	20	20	0	60
Cyber risk management	40	20	10	0	30

4.2. Organisation of the processing and use of big data in Russian financial institutions

The process of processing and storing big data is organised in the vast majority of financial organisations as follows. A unified data storage has been created, the operation and development of which are the responsibility of a separate division (data processing department etc), and each business block has its own specialists or divisions responsible for working with data. At the same time, some financial institutions have a designated unit at the organisational level which is also responsible for the development of all systems and models related to the use of AI; however, in most financial institutions, each business unit's own functional specialists (departments) are responsible for developing models.

In addition to internal sources of big data, financial institutions collect a large set of **external data**. Primarily, the following sources are used (in descending order of the frequency of mention):

83%	data of mobile operators
67%	payment data
58%	geolocation data
33%	social media data
33%	news from the press
25%	satellite imagery
17%	other
17%	only internal data sources are used

All financial institutions indicated the need for additional **data from government agencies**. In particular, banks would like to receive data from the Public Services Portal, the Federal Tax Service (data on taxes and transactions), the Pension Fund of the Russian Federation, the Ministry of Internal Affairs (criminal and administrative cases), the Federal Bailiffs Service (data on enforcement proceedings), the Federal Migration Service, Rosreestr (including information on movable and immovable property owned), the Unified State Register of Legal Entities, the Unified State Register of Individual Entrepreneurs, the housing and communal service system (payments), the civil registry office (information on marital status), the traffic police, the Ministry of Labour (data on employment), courts (claims pending), the Bank of Russia (including payment transactions, faster payment system), passport data and telephone numbers of citizens, registration addresses, individual insurance account and compulsory health insurance numbers, medical data of citizens. For legal entities, banks are also interested in complete data on funds raised by all credit institutions. According to financial institutions, this data will allow them to more correctly assess the financial standing of borrowers and offer them better products and services.

Most financial institutions stated that **depersonalised microdata** is potentially valuable but is not currently used. In particular, banks would like to receive depersonalised microdata from credit history bureaus, which could be used in models for predicting customer behaviour depending on the macroeconomic situation. One of the banks reported that it plans to use depersonalised microdata of telecom operators to predict customer flows in the bank's offices. According to financial institutions, one of the reasons for limited use of depersonalised data is regulatory gaps, in particular, the need to more clearly define the concept of 'depersonalised data' and determine its legal ownership status. Some financial institutions noted that at the moment approaches to data depersonalisation are not regulated by regulatory authorities, as a result of which the transfer of detailed customer data is impossible without a significant loss of data quality; furthermore, there is no unified data transfer standard.

However, the situation with the use of depersonalised microdata may change in the future: in November 2021, President of the Russian Federation Vladimir Putin gave instructions to adopt legislative decisions that will provide Russian AI developers, scientific organisations and businesses with access to state depersonalised data arrays.

4.3. Positive effects of using big data (according to survey data)

According to the financial institutions that took part in the survey, the use of big data has an undoubtedly positive economic effect, but it is difficult to estimate its specific monetary equivalent as big data processing technologies are already an integral part of many business processes, and their financial result is not singled out. The effect is also manifested in the qualitative improvement of business processes. One of the respondent banks estimated the effect of using big data on its profits at about 20%. The exchange estimated annual revenue growth attributed to the use of big data at 3% or more. However, most institutions noted that they expect growth in potential profits from the use of big data in the future.

Among the **main advantages of using big data** for the financial system (both at the organisational level and at the macro level), Russian financial institutions point out the following (in descending order of the frequency of mention):

- improved quality of risk management (100% of respondents)
- optimisation of operating activities and business processes (100%)
- the ability to provide personalised services to a larger number of customers (100%)
- increase in the customer base (75%)
- assistance in the development of new products (67%)
- increased accessibility of financial services (67%)
- assistance in achieving AML/CFT objectives as well as information and economic security objectives (58%).

4.4. Risks of the use of big data by Russian financial institutions (according to survey data)

The risks of the use of big data by **banks** primarily include the following:

- methodological and model risks, including risks associated with the quality and completeness of data and the qualification of employees working with models (92%)
- risks associated with the protection of personal data (67%)
- growing dependence on third-party service providers (cloud solutions, platform and model developers) (58%)
- risks associated with grey areas in regulation (42%)
- risks associated with the occurrence of price discrimination (25%).

Other financial institutions (besides banks) that took part in the survey pointed out primarily methodological and model risks and risks associated with personal data protection as well as an increase in dependence on third-party service providers.

The survey showed that in order to reduce the risks associated with the protection of personal data in most banks (75%) and other financial institutions big data is stored only on their own servers; however, some banks, in addition to their own servers, also use external (cloud) data storage. Those organisations that use external data storages apply special measures to ensure the safety and confidentiality of data as required by law and internal standards, including data depersonalisation, encryption, multi-factor authentication when accessing data etc. Compliance with these measures is monitored and verified by independent information security departments.

To reduce the risks associated with grey zones in regulation, the Bank of Russia is currently developing a number of legislative initiatives. For example, questions concerning the peculiarities of circulation of data at the disposal of financial institutions are being addressed. Also, the regulation of the circulation of depersonalised personal data is being developed together with the Ministry of Digital Development of Russia.

To improve the quality of management of methodological and model risks in financial institutions, the risks of dependence on third-party providers of data and services for its processing and the risks of various forms of discrimination, this report offers a number of steps in the development of principles for managing these risks (see Chapter 5 of this report).

Thus, the Bank of Russia's survey of the largest financial institutions revealed that technologies for working with big data already play a significant role in the activities of financial institutions, and this role will grow. These technologies are used in most key business processes and have a positive effect on financial performance. The main areas of application of big data technologies are credit scoring, risk management and customer analytics. Financial institutions would like to significantly expand the range of external sources of big data through the databases of public authorities. At the same time, there are risks of using big data, including personal data; to restrict these risks, banks apply data protection procedures. Most financial institutions have created a dedicated centralised big data unit.

5. APPROACHES TO REGULATING THE RISKS OF USING BIG DATA TECHNOLOGIES ABROAD AND POSSIBLE MEASURES IN RUSSIA

The financial market shows a lot of interest in using big data processing technologies as they enable financial institutions to better assess their risks and reduce costs and help customers receive additional benefits. Nevertheless, the use of these technologies is fraught with risks, and the regulator's task should be to limit them.

5.1. Foreign regulatory practices

Regulatory approaches to big data technologies, including AI, are at their early stages of development. For example, according to an EBA study, current regulation at this stage sufficiently covers matters related to the use of information technology (including risks associated with data protection, data quality and business continuity and sustainability), security and governance. Further regulatory development may focus on data management and ethical aspects.

Currently, regulatory requirements and guidelines regarding the use of big data processing technologies are often complex and cover the following aspects: ***ensuring non-discrimination of users, ethical use of big data technologies (including in terms of personal data protection) and transparency in the interpretation of model results***.

For example, in January 2019, the **New York Department of Financial Services** (NYDFS) issued a circular with guidelines (requirements) for the use of alternative data by life insurers. The guidelines state that the use of such data must not lead to unlawful discrimination against consumers; insurers must ensure that algorithms and models are consistent with the principles of actuarial valuation, with sufficient justification for any claimed correlation or causation. Moreover, the insurer must disclose to customers the content and source of any external data on the basis of which the insurer made an unfavourable decision.

In July 2021, the **Financial Services Commission of the Republic of Korea** (FSC) published guidance with minimum requirements for financial institutions in terms of the use of AI: the requirements relate to the implementation of necessary internal controls, ensuring proper use of personal financial data, preventing discrimination and ensuring protection of consumer rights. The requirements will become mandatory as soon as all necessary preparatory procedures are completed (the date is not specified).

In 2018, the EU enacted the **General Data Protection Regulation (GDPR)**, which sets out the basic principles of personal data handling, which should also be taken into account when analysing financial data regulation. In particular, the principles laid down in the GDPR provide that processing of personal data be ensured by obtaining targeted, express consent to its use and that data security start with system design (privacy by design). For non-compliance with the GDPR requirements, significant fines are established based on the turnover of companies (up to EUR 20 million or 4% of the company's annual turnover, whichever is higher). This regulation had a significant impact on work with big personalised data sets and analytics based on them.

A number of regulators have also published guidelines for financial institutions to ensure non-discriminatory and ethical use of big data technologies and model transparency. **The European Insurance and Occupational Pensions Authority** (EIOPA) has published its guidelines for managing the use of AI in the insurance sector. The document contains, among other things, the following recommendations: adherence to the principle of fairness and non-discrimination in the application of technologies related to AI, the principle of transparency and explainability of the models used and the need for supervision by employees throughout all stages of the use of AI. It stresses the need to comply with the principles of data management and to ensure the sustainability of the infrastructure and systems associated with AI.

In November 2021, the **EBA** published a consultation paper on the principles of appropriate use of ML technologies in internal rating models (IRBs) to calculate regulatory capital for covering credit risk. In particular, the document includes the following principles: the need for an appropriate level of personnel qualification for the development and validation of ML models; providing senior management with documentation on how ML models work. It is recommended to avoid over-complicating the model. Moreover, the model must enable correct interpretation and have explainable results and reliable model validation.

The **German Federal Financial Supervisory Authority** (BaFin) has published supervisory guidelines for responsible use of big data and AI by financial institutions. These principles provide preliminary ideas for the subsequent development of minimum supervisory requirements. The document presents general principles of using algorithms in decision-making (the need to delimit responsibilities within an organisation and develop a risk management system adapted to the use of algorithms in decision-making; ensuring unbiased results of decision-making based on algorithms and the absence of customer discrimination). Moreover, the paper addresses specific principles that must be taken into account when implementing algorithms based on big data and AI (the need to develop a strategy for checking the quality of data in algorithms and compliance with personal data protection requirements; development of documentation for verification of algorithms, either by the company itself or by auditors and supervisors; effective procedures for algorithm validation and data sampling) as well as specific principles that need to be considered when applying and interpreting algorithms (for example, the organisation's employees should be involved in interpreting algorithm results when making decisions; an action plan should be made for the event of failures in decision-making processes based on algorithms).

The Monetary Authority of Singapore (MAS) has developed voluntary guidelines aimed at ensuring integrity, ethics, accountability and transparency in the use of AI and data analysis (AIDA) in the financial sector. In terms of ensuring good faith, the authors stressed the need to ensure protection against systematic infringement of the interests of individuals and/or groups of individuals by making decisions using AI technologies, regularly checking models for correct operation and eliminating systematic errors. Models must function in accordance with set objectives. Ethics means that these technologies will be used in accordance with ethical standards, values and codes of conduct established by the company. Decisions made on the basis of AIDA should meet at least the same ethical standards as decisions made by humans. The responsibility principles address the need for companies to be held accountable for the application of both their in-house and outsourced models. Transparency is understood as the need to provide data subjects, upon their request, with clear explanations regarding the possible consequences of decisions made on the basis of AI that affect their interests as well as with respect to what kind of data about specific subjects is used in making such decisions.

The **Office of the Privacy Commissioner for Personal Data of Hong Kong** developed an Ethical Accountability Framework (EAF), which generally suggested the promotion of a culture of ethical data management and reduction of risks associated with the use of personal data, including within the framework of systems based on the use of big data. In 2019, the **Hong Kong Monetary Authority** (HKMA) issued a circular urging financial institutions to unconditionally adhere to EAF principles and a circular aimed at strengthening financial consumer protection when using big data and AI technology. The circular contains a number of guidelines in the following areas: governance and accountability of big data and AI systems, fair application of big data and AI systems, transparency and disclosure of information, information protection.

A report with similar principles was also published by the **De Nederlandsche Bank** in 2019 (with a focus on six principles of using AI in financial services: soundness, accountability, fairness, ethics, skills and transparency (SAFEST).

On 1 November 2021, a **law on personal data protection** took effect in **China**: internet services and applications are now required to provide users with information processing options not requiring

their personal data or offer an option to refuse to provide such information. The new law also obliges internet resources to obtain user consent to the processing of their biometric data, health information, financial accounts and geolocation. This law will have a great impact on bigtechs (such as Alibaba or Tencent), which accumulate significant volumes of information.

Individual regulators, in addition to addressing issues of customer non-discrimination and ethical and responsible use of big data processing technologies, also consider ***issues of obtaining information for regulatory purposes***. For example, in June 2020, the **People's Bank of China** announced its intention to create a single information base for the financial sector based on big data processing technologies. This database should become a single centre for managing the collection, storage and analysis of financial information. Furthermore, in February 2021, the Chinese regulator published a report justifying the need to introduce ***requirements for large internet platforms to provide full credit information to the regulator***. Data on millions of users of internet platforms are used by large technology companies (Ant Group (Alibaba), WeBank (Tencent) etc) to provide loans (while the loan funds themselves are provided by banks, with internet platforms only acting as intermediaries in such transactions).

In June 2021, the **Financial Services Commission** of the Republic of Korea (FSC) announced its plans for the development of big data analytics in the financial sector. The said measures include plans for increasing the number of companies specialising in **data convergence**, partial lifting of restrictions on data convergence and its subsequent use by the same company in the absence of conflicts of interest (according to the current legislation, processed data may be used by third parties only), removing the restriction on the use of partial data samples and expanding access to data convergence and use for SMEs that do not have their own data samples but express a wish to use data provided by other companies (this will reduce the administrative burden on companies owning original data).

In 2019, the Organisation for Economic Co-operation and Development (OECD) published a recommendation which is essentially the first intergovernmental standard on AI[1]. The document includes five principles of using AI: promoting inclusive growth, sustainable development and well-being; human-centred values and fairness; transparency and explainability; robustness, security and safety; accountability. Moreover, adequate operation of AI should be ensured in accordance with the above principles. In turn, governments are encouraged to invest in AI R&D to stimulate the development of digital ecosystems for AI; create conditions favourable for the introduction of these technologies; stimulate training and increase the competence of employees; operate in the international arena and promote international cooperation between different sectors of the economy on the exchange of information and knowledge, development of standards and implementation of principles.

5.2. Possible steps in the field of regulation of the use of big data technologies in Russia to reduce risks to financial stability

Currently, Russian regulation covers only some aspects of the use of big data.

For example, Federal Law No. 152-FZ, dated 27 July 2006, 'On Personal Data' was adopted to ensure personal data protection. The objective of this Federal Law is to ensure protection of the rights and freedoms of a person and citizen when their personal data is processed, including the protection of rights to privacy, personal and family secrets.

Bank of Russia Regulation No. 716-P, dated 8 April 2020, 'On Requirements for the Operational Risk Management System in a Credit Institution and a Banking Group' ('Regulation No. 716-P'), Bank of Russia Regulation No. 483-P, dated 6 August 2015, 'On the Procedure for Calculating the Amount of Credit Risk Based on Internal Ratings' ('Regulation No. 483-P') and other regulatory documents

[1] *42 countries have acceded to the agreement.*

of the Bank of Russia establish requirements for the management of model risk by banks within the framework of the operational risk management system, in the ICAAP and when applying the IRB approach for the purpose of assessing capital adequacy.

In addition, Regulation No. 716-P specifies the Bank of Russia requirements for the reliability and continuity of information systems and assurance of the quality of data in information systems within the framework of the operational risk management system, requirements for ensuring information security as well as Bank of Russia requirements related to the interaction of credit institutions with third-party service providers with the purpose of complying with information security requirements and to outsourcing information system elements and data of credit institutions.

Control (monitoring) of risks associated with the use of big data and its assessment should be implemented within the scope of the credit institution's risk management system, provided that the credit institution's internal control system functions effectively.

To address additional risks of using big data presented in this report, given the growing role of big data in the banking business, the **Bank of Russia proposes discussing the following measures**.

Together with banks and other financial institutions, develop a set of principles for proper use of big data, which could be further implemented in the form of standards for self-regulatory organisations and professional associations of financial market participants. In the part where the subject area of these principles intersects with the current regulation, these principles may serve to promote the existing regulation. The experience of applying this set of principles and the conclusions obtained regarding their relevance and efficiency can be further used by the Bank of Russia to develop its recommendations for the proper use of big data.

This set of principles may also be viewed as the development of a National Code of Ethics for Artificial Intelligence[2] in respect of financial institutions.

In particular, it is proposed to include the following clauses in this **set of principles**:

1. A financial institution should develop a data management strategy (or another top-level document) which would, in particular, regulate matters of big data management, including the procedure for collecting and storing it, ensuring the completeness and quality of data, the responsible persons (departments), the responsibilities of the organisation's management and the procedure for conducting internal audit inspections, taking into account, inter alia, the requirements of Regulation No. 716-P and Appendix 3 to Regulation No. 483-P.
2. To limit model risks and quality risks of big data, institutions should develop a system for ensuring the completeness and quality of data to prevent situations where models developed on the basis of big data provide biased or incorrect results or results leading to discrimination due to incomplete data or improper quality of the models themselves. A system should be developed for regular updates of big data according to the nature of such data. Models should be calibrated and independently validated on a regular basis. The models themselves must be transparent, interpretable and available for independent review by external auditors and the regulator. The institution's internal control system should review the data and model risk management system.
3. To limit risks associated with the ethical use of big data, special **ethical principles** should be developed that can be established as a self-regulatory organisations standard.

   In particular, these ethical principles should cover the use of big data in marketing (CRM) and pricing, especially data obtained from external sources, and the results of models developed using this data. Among other things, aspects related to safety, confidentiality and ethical use should be regulated in respect of external big data, such as data from social media, internet searches, mobile

[2] *The National Code of Ethics in the Field of Artificial Intelligence signed on 26 October 2021 in Moscow by leading technology companies within the framework of the 1st International Forum on the Ethics of Artificial Intelligence was developed as part of the implementation of the National Strategy for the Development of Artificial Intelligence up to 2030 approved by Decree of the President of the Russian Federation No. 490, dated 10 October 2019. The Code was developed with the participation of the Alliance in the Field of Artificial Intelligence. The provisions of the Code can be expanded and made more specific for individual economic sectors, taking into account their specific features (Clause 2 of the Code 'Permissibility of Sectoral Codes').*

operators, payment data, geolocation data and other data (which is not depersonalised). Ethical use means, in particular, ***the absence of any discrimination, be it price discrimination or other kinds, including gender, racial, national etc, the absence of any restrictions on the freedom of choice of financial services*** and the observance of human rights and freedoms (including the absence of abuse of private user information posted on the internet). In addition, these principles should regulate the use of big data in direct marketing (such as cold calls, distribution of so-called 'pre-approved offers' etc).

4. To manage the risk of dependence on third-party providers of big data[3] or services for its processing, a financial institution's internal documents should provide for appropriate approaches, in particular, for monitoring the share of each individual provider of big data (as well as models based on it and storage and processing services), stress testing of operational risk in the event of disruptions in the work of suppliers and the procedure for restoring operations in the event of disruptions. The country aspect of risk should also be taken into account by diversifying foreign providers and using the services of national suppliers where possible. Moreover, institutions should control the quality of data and models received from external providers and stipulate the liability of providers for data quality. In any case, institutions should be ultimately responsible for the application of both their in-house and outsourced models.

   In the future, the Bank of Russia is expected to consider establishing requirements for outsourcing risk management within the operational risk management system, as well as introducing requirements for prior notification of the regulatory body about the outsourcing of significant functions for the purpose of identifying risks of third-party service providers, including concentration risks and warnings of the emergence of new systemic risks in connection therewith.

   In general, we believe it expedient to define approaches to the management of risks associated with the use of big data (for example, model risk, information system risk, information security risk and outsourcing risk), taking into account the requirements for operational risk management.

[3] *This does not apply to credit history bureaus as data providers.*

DISCUSSION QUESTIONS

1. Do you see the need to develop a big data strategy at the institutional level? What questions should such a strategy include?
2. What approaches does your institution use (plan to use) to manage model and data quality risks with respect to big data?
3. What approaches does your institution use (plan to use) to manage the risks of discrimination (including non-price discrimination) in making decisions using big data?
4. How does your institution manage the risk of dependence on third-party service providers for big data and models built on its basis? Does your institution see an increased level of concentration on any providers? What is the maximum share of an individual provider of big data and big data services in the core business areas of your institution?
5. Do you support the creation of the abovementioned set of principles for proper use of big data by financial institutions? What key points and principles does your institution consider appropriate to include in this set?
6. What other big data regulations does your institution consider most relevant and appropriate for development and implementation?
7. What information would your institution like to receive in depersonalised form to improve its use of big data?
8. Is it possible, in your institution's opinion, to establish the principles of working with big data, including ethical principles, as standards of self-regulatory organisations and professional associations of financial market players?