READING

43

Fintech in Investment Management

by Barbara J. Mack and Robert Kissell, PhD

Barbara J. Mack is at Pingry Hill Enterprises, Inc. (USA). Robert Kissell, PhD, is at Kissell Research Group (USA).

LEARNING OUTCOMES

Mastery	*The candidate should be able to*
☐	**a.** describe "fintech;"
☐	**b.** describe Big Data, artificial intelligence, and machine learning;
☐	**c.** describe fintech applications to investment management;
☐	**d.** describe financial applications of distributed ledger technology.

INTRODUCTION

1

The meeting of finance and technology, commonly known as *fintech*, is changing the landscape of investment management. Advancements include the use of Big Data, artificial intelligence, and machine learning to evaluate investment opportunities, optimize portfolios, and mitigate risks. These developments are affecting not only quantitative asset managers but also fundamental asset managers who make use of these tools and technologies to engage in hybrid forms of investment decision making.

Investment advisory services are undergoing changes with the growth of automated wealth advisers or "robo-advisers." Robo-advisers may assist investors without the intervention of a human adviser, or they may be used in combination with a human adviser. The desired outcome is the ability to provide tailored, actionable advice to investors with greater ease of access and at lower cost.

In the area of financial record keeping, blockchain and distributed ledger technology (DLT) are creating new ways to record, track, and store transactions for financial assets. An early example of this trend is the cryptocurrency bitcoin, but the technology is being considered in a broader set of applications.

This reading is divided into seven main sections, which together define fintech and outline some of its key areas of impact in the field of investment management. Section 2 explains the concept of and areas of fintech. Sections 3 and 4 discuss Big

Data, artificial intelligence, and machine learning. Section 5 discusses data science, and Section 6 provides applications of fintech to investment management. Section 7 examines DLT. A summary of key points completes the reading.

2 WHAT IS FINTECH?

In its broadest sense, the term "fintech" generally refers to technology-driven innovation occurring in the financial services industry. For the purposes of this reading, **fintech** refers to technological innovation in the design and delivery of financial services and products. Note, however, that in common usage, fintech can also refer to companies (often new, startup companies) involved in developing the new technologies and their applications, as well as the business sector that comprises such companies. Many of these innovations are challenging the traditional business models of incumbent financial services providers.

Early forms of fintech included data processing and the automation of routine tasks. Then followed systems that provided execution of decisions according to specified rules and instructions. Fintech has since advanced into decision-making applications based on complex machine-learning logic, where computer programs are able to "learn" how to complete tasks over time. In some applications, advanced computer systems are performing tasks at levels far surpassing human capabilities. Fintech has changed the financial services industry in many ways, giving rise to new systems for investment advice, financial planning, business lending, and payments.

Whereas fintech covers a broad range of services and applications, areas of fintech development that are more directly relevant to the investment industry include the following:

- **Analysis of large datasets.** In addition to growing amounts of traditional data, such as security prices, corporate financial statements, and economic indicators, massive amounts of alternative data generated from non-traditional data sources, such as social media and sensor networks, can now be integrated into a portfolio manager's investment decision-making process and used to help generate alpha and reduce losses.
- **Analytical tools.** For extremely large datasets, techniques involving **artificial intelligence** (AI)—computer systems capable of performing tasks that previously required human intelligence—may be better suited to identify complex, non-linear relationships than traditional quantitative methods and statistical analysis. Advances in AI-based techniques are enabling different data analysis approaches. For example, analysts are turning to artificial intelligence to sort through the enormous amounts of data from company filings, annual reports, and earnings calls to determine which data are most important and to help uncover trends and generate insights relating to human sentiment and behavior.
- **Automated trading.** Executing investment decisions through computer algorithms or automated trading applications may provide a number of benefits to investors, including more efficient trading, lower transaction costs, anonymity, and greater access to market liquidity.

- **Automated advice. Robo-advisers** or automated personal wealth management services provide investment services to a larger number of retail investors at lower cost than traditional adviser models can provide.
- **Financial record keeping.** New technology, such as DLT, may provide secure ways to track ownership of financial assets on a peer-to-peer (P2P) basis. By allowing P2P interactions—in which individuals or firms transact directly with each other without mediation by a third party—DLT reduces the need for financial intermediaries.

Drivers underlying fintech development in these areas include extremely rapid growth in data—including their quantity, types, sources, and quality—and technological advances that enable the capture and extraction of information from them. The data explosion is addressed in Section 3, and selected technological advances and data science are addressed in Sections 4 and 5, respectively.

BIG DATA

3

As noted, datasets are growing rapidly in terms of the size and diversity of data types that are available for analysis. The term **Big Data** has been in use since the late 1990s and refers to the vast amount of data being generated by industry, governments, individuals, and electronic devices. Big Data includes data generated from traditional sources—such as stock exchanges, companies, and governments—as well as non-traditional data types, also known as **alternative data**, arising from the use of electronic devices, social media, sensor networks, and company exhaust (data generated in the normal course of doing business).

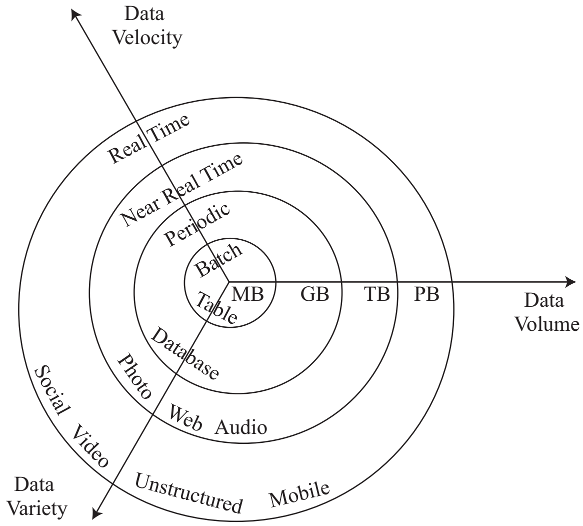
Traditional data sources include corporate data in the form of annual reports, regulatory filings, sales and earnings figures, and conference calls with analysts. Traditional data also include data that are generated in the financial markets, including trade prices and volumes. Because the world has become increasingly connected, we can now obtain data from a wide range of devices, including smart phones, cameras, microphones, radio-frequency identification (RFID) readers, wireless sensors, and satellites that are now in use all over the world. As the internet and the presence of such networked devices have grown, the use of non-traditional data sources, or alternative data sources—including social media (posts, tweets, and blogs), email and text communications, web traffic, online news sites, and other electronic information sources—has risen.

The term *Big Data* typically refers to datasets having the following characteristics:

- **Volume**: The amount of data collected in files, records, and tables is very large, representing many millions, or even billions, of data points.
- **Velocity**: The speed with which the data are communicated is extremely great. Real-time or near-real-time data have become the norm in many areas.
- **Variety**: The data are collected from many different sources and in a variety of formats, including structured data (e.g., SQL tables or CSV files), semi-structured data (e.g., HTML code), and unstructured data (e.g., video messages).

Features relating to Big Data's volume, velocity, and variety are shown in Exhibit 1.

Exhibit 1 Big Data Characteristics: Volume, Velocity, and Variety



Data	Volume Key	Bytes of Information
MB	Megabyte	One Million
GB	Gigabyte	One Billion
TB	Terabyte	One Trillion
PB	Petabyte	One Quadrillion

Source: http://whatis.techtarget.com/definition/3Vs.

Exhibit 1 shows that data volumes are growing from megabytes (MB) and gigabytes (GB) to far larger sizes, such as terabytes (TB) and petabytes (PB), as more data are being generated, captured, and stored. At the same time, more data, traditional and non-traditional, are available on a real-time or near-real-time basis with far greater variety in data types than ever before.

Big Data may be structured, semi-structured, or unstructured data. Structured data items can be organized in tables and are commonly stored in a database where each field represents the same type of information. Unstructured data may be disparate, unorganized data that cannot be represented in tabular form. Unstructured data, such as those generated by social media, email, text messages, voice recordings, pictures, blogs, scanners, and sensors, often require different, specialized applications or custom programs before they can be useful to investment professionals. For example, in order to analyze data contained in emails or texts, specially developed or customized computer code may be required to first process these files. Semi-structured data may have attributes of both structured and unstructured data.

3.1 Sources of Big Data

Big Data, therefore, encompasses data generated by

- financial markets (e.g., equity, fixed income, futures, options, and other derivatives),
- businesses (e.g., corporate financials, commercial transactions, and credit card purchases),
- governments (e.g., trade, economic, employment, and payroll data),

- individuals (e.g., credit card purchases, product reviews, internet search logs, and social media posts),
- sensors (e.g., satellite imagery, shipping cargo information, and traffic patterns), and, in particular,
- the Internet of Things, or IoT (e.g., data generated by “smart” buildings, where the building is providing a steady stream of information about climate control, energy consumption, security, and other operational details).

In gathering business intelligence, historically, analysts have tended to draw on traditional data sources, employing statistical methods to measure performance, predict future growth, and analyze sector and market trends. In contrast, the analysis of Big Data incorporates the use of alternative data sources.

From retail sales data to social media sentiment to satellite imagery that may reveal information about agriculture, shipping, and oil rigs, alternative datasets may provide additional insights about consumer behavior, firm performance, trends, and other factors important for investment-related activities. Such information is having a significant effect on the way that professional investors, particularly quantitative investors, approach financial analysis and decision-making processes.

There are three main sources of alternative data:

- data generated by individuals,
- data generated by business processes, and
- data generated by sensors.

Data generated by individuals are often produced in text, video, photo, and audio formats and may also be generated through such means as website clicks or time spent on a webpage. This type of data tends to be unstructured. The volume of this type of data is growing dramatically as people participate in greater numbers and more frequently in online activities, such as social media and e-commerce, including online reviews of products, services, and entire companies, and as they make personal data available through web searches, email, and other electronic trails.

Business process data include information flows from corporations and other public entities. These data tend to be structured data and include direct sales information, such as credit card data, as well as corporate exhaust. Corporate exhaust includes corporate supply chain information, banking records, and retail point-of-sale scanner data. Business process data can be leading or real-time indicators of business performance, whereas traditional corporate metrics may be reported only on a quarterly or even yearly basis and are typically lagging indicators of performance.

Sensor data are collected from such devices as smart phones, cameras, RFID chips, and satellites that are usually connected to computers via wireless networks. Sensor data can be unstructured, and the volume of data is many orders of magnitude greater than that of individual or business process datastreams. This form of data is growing exponentially because microprocessors and networking technology are increasingly present in a wide array of personal and commercial electronic devices. Extended to office buildings, homes, vehicles, and many other physical forms, this culminates in a network arrangement, known as the **Internet of Things**, that is formed by the vast array of physical devices, home appliances, smart buildings, vehicles, and other items that are embedded with electronics, sensors, software, and network connections that enable the objects in the system to interact and share information.

Exhibit 2 shows a classification of alternative data sources and includes examples for each.

Exhibit 2 Classification of Alternative Data Sources

Individuals	Business Processes	Sensors
Social media	Transaction data	Satellites
News, reviews	Corporate data	Geolocation
Web searches, personal data		Internet of Things
		Other sensors

In the search to identify new factors that may affect security prices, enhance asset selection, improve trade execution, and uncover trends, alternative data are being used to support data-driven investment models and decisions. As interest in alternative data has risen, there has been a growth in the number of specialized firms that collect, aggregate, and sell alternative datasets.

While the marketplace for alternative data is expanding, investment professionals should understand potential legal and ethical issues related to information that is not in the public domain. For example, the scraping of web data could potentially capture personal information that is protected by regulations or that may have been published or provided without the explicit knowledge and consent of the individuals involved. Best practices are still in development in many jurisdictions, and because of varying approaches taken by national regulators, there may be conflicting forms of guidance.

3.2 Big Data Challenges

Big Data poses several challenges when it is used in investment analysis, including the quality, volume, and appropriateness of the data. Key issues revolve around the following questions, among others: Does the dataset have selection bias, missing data, or data outliers? Is the volume of collected data sufficient? Is the dataset well suited for the type of analysis? In most instances, the data must be sourced, cleansed, and organized before analysis can occur. This process can be extremely difficult with alternative data owing to the unstructured characteristics of the data involved, which are more often qualitative (e.g., texts, photos, and videos) than quantitative in nature.

Given the size and complexity of alternative datasets, traditional analytical methods cannot always be used to interpret and evaluate these datasets. To address this challenge, artificial intelligence and machine learning techniques have emerged that support work on such large and complex sources of information.

4 ADVANCED ANALYTICAL TOOLS: ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Artificial intelligence computer systems are capable of performing tasks that have traditionally required human intelligence. AI technology has enabled the development of computer systems that exhibit cognitive and decision-making ability comparable or superior to that of human beings.

An early example of AI was the “expert system,” a type of computer programming that attempted to simulate the knowledge base and analytical abilities of human experts in specific problem-solving contexts. This was often accomplished through the use of “if-then” rules. By the late 1990s, faster networks and more powerful processors enabled AI to be deployed in logistics, data mining, financial analysis, medical diagnosis, and other areas. Since the 1980s, financial institutions have made use of AI—particularly, **neural networks**, programming based on how our brain learns and processes information—to detect abnormal charges or claims in credit card fraud detection systems.

Machine learning (ML) is a technology that has grown out of the wider AI field. ML algorithms are computer programs that are able to "learn" how to complete tasks, improving their performance over time with experience. As it is currently used in the investing context, ML requires massive amounts of data for "training," so although some ML techniques have existed for years, insufficient data have historically limited broader application. Previously, these algorithms lacked access to the large amounts of data needed to model relationships successfully. The growth in Big Data has provided ML algorithms, such as neural networks, with sufficient data to improve modeling and predictive accuracy, and greater use of ML techniques is now possible.

In ML, the computer algorithm is given "inputs" (a set of variables or datasets) and may be given "outputs" (the target data). The algorithm "learns" from the data provided how best to model inputs to outputs (if provided) or how to identify or describe underlying data structure if no outputs are given. Training occurs as the algorithm identifies relationships in the data and uses that information to refine its learning process.

ML involves splitting the dataset into a training dataset and validation dataset (evaluation dataset). The training dataset allows the algorithm to identify relationships between inputs and outputs based on historical patterns in the data. These relationships are then tested on the validation dataset. Once an algorithm has mastered the training and validation datasets, the ML model can be used to predict outcomes based on other datasets.

ML still requires human judgement in understanding the underlying data and selecting the appropriate techniques for data analysis. Before they can be used, the data must be clean and free of biases and spurious data. As noted, ML models also require sufficiently large amounts of data and may not perform well where there may not be enough available data to train and validate the model.

Analysts must also be cognizant of errors that may arise from **overfitting** the data, because models that overfit the data may discover "false" relationships or "unsubstantiated" patterns that will lead to prediction errors and incorrect output forecasts. Overfitting occurs when the ML model learns the input and target dataset too precisely. In such cases, the model has been "over-trained" on the data and treats noise in the data as true parameters. An ML model that has been overfitted is not able to accurately predict outcomes using a different dataset and may be too complex. When a model has been underfitted, the ML model treats true parameters as if they are noise and is not able to recognize relationships within the training data. In such cases, the model may be too simplistic. Underfitted models will typically fail to fully discover patterns that underlie the data.

In addition, since they are not explicitly programmed, ML techniques can appear to be opaque or "black box" approaches, which arrive at outcomes that may not be entirely understood or explainable.

4.1 Types of Machine Learning

ML approaches can help identify relationships between variables, detect patterns or trends, and create structure from data, including data classification. The main types of machine learning approaches include supervised and unsupervised learning.

In **supervised learning**, computers learn to model relationships based on labeled training data. In supervised learning, inputs and outputs are labeled, or identified, for the algorithm. After learning how best to model relationships for the labeled data, the trained algorithms are used to model or predict outcomes for new datasets. Trying to identify the best signal, or variable, to forecast future returns on a stock or trying to predict whether local stock market performance will be up, down, or flat during the next business day are problems that may be approached using supervised learning techniques.

In **unsupervised learning**, computers are not given labeled data but instead are given only data from which the algorithm seeks to describe the data and their structure. Trying to group companies into peer groups based on their characteristics rather than using standard sector or country groupings is a problem that may be approached using unsupervised learning techniques.

Underlying AI advances have been key developments relating to neural networks. In **deep learning**, (or **deep learning nets**), computers use neural networks, often with many hidden layers, to perform multistage, non-linear data processing to identify patterns. Deep learning may use supervised or unsupervised machine learning approaches. By taking a layered or multistage approach to data analysis, deep learning develops an understanding of simple concepts that informs analysis of more complex concepts. Neural networks have existed since 1958 and have been used for many applications, such as forecasting and pattern recognition, since the early 1990s. Improvements in the algorithms underlying neural networks are providing more accurate models that better incorporate and learn from data. As a result, these algorithms are now far better at such activities as image, pattern, and speech recognition. In many cases, the advanced algorithms require less computing power than the earlier neural networks, and their improved solution enables analysts to discover insights and identify relationships that were previously too difficult or too time consuming to uncover.

Advances in Artificial Intelligence outside Finance

Non-finance-related AI breakthroughs include victories in the general knowledge game-show Jeopardy (by IBM's Watson in 2011) and in the ancient Chinese board game Go (by Google's DeepMind in 2016). Not only is AI providing solutions where there is perfect information (all players have equal access to the same information), such as checkers, chess, and Go, but AI is also providing insight in cases where information may be imperfect and players have hidden information; AI successes at the game of poker (by DeepStack) are an example. AI has also been behind the rise of virtual assistants, such as Siri (from Apple), Google's Translate app, and Amazon's product recommendation engine.

The ability to analyze Big Data using ML techniques, alongside more traditional statistical methods, represents a significant development in investment research, supported by the presence of greater data availability and advances in the algorithms themselves. Improvements in computing power and software processing speeds and falling storage costs have further supported this evolution.

ML techniques are being used for Big Data analysis to help predict trends or market events, such as the likelihood of a successful merger or an outcome to a political election. Image recognition algorithms can now analyze data from satellite-imaging systems to provide intelligence on the number of consumers in retail store parking lots, shipping activity and manufacturing facilities, and yields on agricultural crops, to name just a few examples.

Such information may provide insights into individual firms or at national or global levels and may be used as inputs into valuation or economic models.

DATA SCIENCE: EXTRACTING INFORMATION FROM BIG DATA

5

Data science can be defined as an interdisciplinary field that harnesses advances in computer science (including machine learning), statistics, and other disciplines for the purpose of extracting information from Big Data (or data in general). Companies rely on the expertise of data scientists/analysts to extract information and insights from Big Data for a wide variety of business and investment purposes.

An important consideration for the data scientist is the structure of the data. As noted in the discussion on Big Data, because of their unstructured nature, alternative data often require specialized treatment before they can be used for analysis.

5.1 Data Processing Methods

To help determine the best data management technique needed for Big Data analysis, data scientists use various data processing methods, including capture, curation, storage, search, and transfer.

- Capture—Data capture refers to how the data are collected and transformed into a format that can be used by the analytical process. Low-latency systems—systems that operate on networks that communicate high volumes of data with minimal delay (latency)—are essential for automated trading applications that make decisions based on real-time prices and market events. In contrast, high-latency systems do not require access to real-time data and calculations.
- Curation—Data curation refers to the process of ensuring data quality and accuracy through a data cleaning exercise. This process consists of reviewing all data to detect and uncover data errors—bad or inaccurate data—and making adjustments for missing data when appropriate.
- Storage—Data storage refers to how the data will be recorded, archived, and accessed and the underlying database design. An important consideration for data storage is whether the data are structured or unstructured and whether analytical needs require low-latency solutions.
- Search—Search refers to how to query data. Big Data has created the need for advanced applications capable of examining and reviewing large quantities of data to locate requested data content.
- Transfer—Transfer refers to how the data will move from the underlying data source or storage location to the underlying analytical tool. This could be through a direct data feed, such as a stock exchange's price feed.

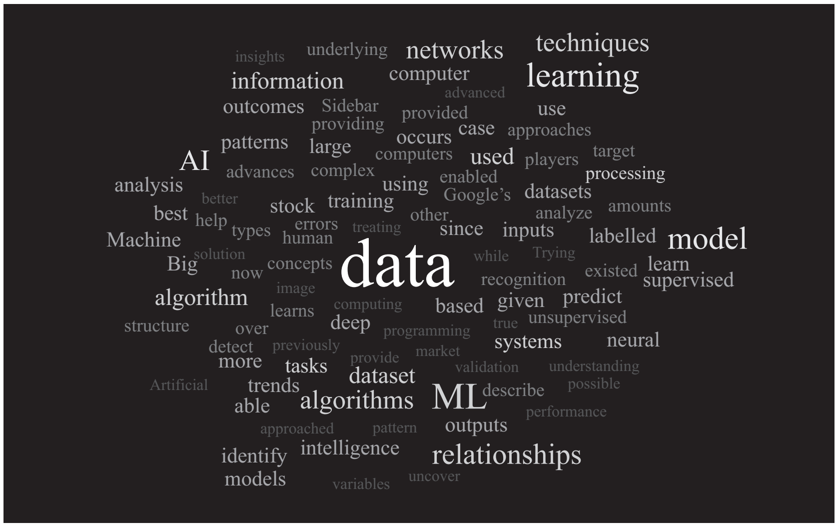
5.2 Data Visualization

Data visualization is an important tool for understanding Big Data. Visualization refers to how the data will be formatted, displayed, and summarized in graphical form. Traditional structured data can be visualized using tables, charts, and trends, whereas non-traditional unstructured data require new techniques of data visualization. These visualization tools include, for example, interactive three-dimensional (3D) graphics, where users can focus in on specified data ranges and rotate the data across 3D axes to help identify trends and uncover relationships. Multidimensional data analysis consisting of more than three variables requires additional data visualization techniques—for example, adding color, shapes, and sizes to the 3D charts. Further, a wide variety of solutions exists to reflect the structure of the data through the geometry of the visualization, with interactive graphics allowing for especially rich possibilities. Examples include heat maps, tree diagrams, and network graphs.

Another valuable Big Data visualization technique that is applicable to textual data is a "tag cloud," where words are sized and displayed on the basis of the frequency of the word in the data file. For example, words that appear more often are shown with a larger font, and words that appear less often are shown with a smaller font. A "mind map" is another data visualization technique; it is a variation of the tag cloud, but rather than displaying the frequency of words, a mind map shows how different concepts are related to each other.

Exhibit 3 shows an example of a "tag cloud" based on a section of this reading. The more frequently a word is found within the text, the larger it becomes in the tag cloud. As shown in the tag cloud, the words appearing most frequently in the section include "data," "ML," "learning," "AI," "techniques," "model," and "relationships."

Exhibit 3 Data Visualization Tag Cloud: Section 4, Advanced Analytical Tools



Source: https://worditout.com/word-cloud/create.

Programming Languages and Databases

Some of the more common programming languages used in data science include the following:

- **Python:** Python is an open source, free programming language that does not require an in-depth understanding of computer programming. Python allows individuals with little or no programming experience to develop computer applications for advanced analytical use and is the basis for many fintech applications.
- **R:** R is an open source, free programming language traditionally used for statistical analysis. R has mathematical packages for statistical analysis, machine learning, optimization, econometrics, and financial analysis.

- **Java:** Java is a programming language that can run on different computers, servers, and operating systems. Java is the underlying program language used in many internet applications.
- **C/C++:** C/C++ is a specialized programming language that provides the ability to optimize source code to achieve superior calculation speed and processing performance. C/C++ is used in applications for algorithmic and high-frequency trading.
- **Excel VBA:** Excel VBA helps bridge the gap between programming and manual data processing by allowing users to run macros to automate tasks, such as updating data tables and formulas, running data queries and collecting data from different web locations, and performing calculations. Excel VBA allows users to develop customized reports and analyses that rely on data that are updated from different applications and databases.

Some of the more common databases in use include the following:

- **SQL:** SQL is a database for structured data where the data can be stored in tables with rows and columns. SQL databases need to be run on a server that is accessed by users.
- **SQLite:** SQLite is a database for structured data. SQLite databases are embedded into the program and do not need to be run on a server. It is the most common database for mobile apps that require access to data.
- **NoSQL:** NoSQL is a database used for unstructured data where the data cannot be summarized in traditional tables with rows and columns.

SELECTED APPLICATIONS OF FINTECH TO INVESTMENT MANAGEMENT

6

Fintech is being used in numerous areas of investment management. Applications for investment management include text analytics and natural language processing, robo-advisory services, risk analysis, and algorithmic trading.

6.1 Text Analytics and Natural Language Processing

Text analytics involves the use of computer programs to analyze and derive meaning typically from large, unstructured text- or voice-based datasets, such as company filings, written reports, quarterly earnings calls, social media, email, internet postings, and surveys. Text analytics includes using computer programs to perform automated information retrieval from different, unrelated sources in order to aid the decision-making process. More analytical usage includes lexical analysis, or the analysis of word frequency in a document and pattern recognition based on key words and phrases. Text analytics may be used in predictive analysis to help identify indicators of future performance, such as consumer sentiment.

Natural language processing (NLP) is a field of research at the intersection of computer science, artificial intelligence, and linguistics that focuses on developing computer programs to analyze and interpret human language. Within the larger field of text analytics, NLP is an important application. Automated tasks using NLP include translation, speech recognition, text mining, sentiment analysis, and topic analysis. NLP may also be employed in compliance functions to review employee voice and electronic communications for adherence to company or regulatory policy, inappropriate conduct, or fraud or for ensuring private or customer information is kept confidential.

Consider that all the public corporations worldwide generate millions of pages of annual reports and tens of thousands of hours of earnings calls each year. This is more information than any individual analyst or team of researchers can assess. NLP, especially when aided by ML algorithms, can analyze annual reports, call transcripts, news articles, social media posts, and other text- and audio-based data to identify trends in shorter timespans and with greater scale and accuracy than is humanly possible.

For example, NLP may be used to monitor analyst commentary to aid investment decision making. Financial analysts may generate earnings-per-share (EPS) forecasts reflecting their views on a company's near-term prospects. Focusing on forecasted EPS numbers could mean investors miss subtleties contained in an analyst's written research report. Since analysts tend not to change their buy, hold, and sell recommendations for a company frequently, they may instead offer nuanced commentary without making a change in their investment recommendation. After analyzing analyst commentary, NLP can assign sentiment ratings ranging from very negative to very positive for each. NLP can, therefore, be used to detect, monitor, and tag shifts in sentiment, potentially ahead of an analyst's recommendation change. Machine capabilities enable this analysis to scale across thousands of companies worldwide, performing work previously done by humans.

Similarly, communications and transcripts from policymakers, such as the European Central Bank or the US Federal Reserve, offer an opportunity for NLP-based analysis, because officials at these institutions may send subtle messages through their choice of topics, words, and inferred tone. NLP can help analyze nuances within text to provide insights around trending or waning topics of interest, such as interest rate policy, aggregate output, or inflation expectations.

Models using NLP analysis may incorporate non-traditional information to evaluate what people are saying—via their preferences, opinions, likes, or dislikes—in an attempt to identify trends and short-term indicators about a company, a stock, or an economic event that might have a bearing on future performance. Past research has evaluated the predictive power of Twitter sentiment regarding IPO performance, for example.[1] The effect of positive and negative news sentiment on stock returns has also been researched.[2]

6.2 Robo-Advisory Services

Since their emergence in 2008, a number of startup firms, as well as large asset managers, have introduced robo-advisory services, which provide investment solutions through online platforms, reducing the need for direct interaction with financial advisers.

As robo-advisers have been incorporated into the investment landscape, they have drawn the attention of regulatory authorities. In the United States, robo-advisers must be established as registered investment advisers, and they are regulated by the Securities and Exchange Commission. In the United Kingdom, they are regulated by the Financial Conduct Authority. In Australia, all financial advisers must obtain an Australian Financial Services license, with guidance on digital advisers coming from the Australian Securities and Investments Commission. Robo-advisers are also on the rise in parts of Asia and the rest of the world. Although regulatory conditions vary, robo-advisers are likely to be held to a similar level of scrutiny and code of conduct as other investment professionals in the given region.

---

1 Jim Kyung-Soo Liew and Garrett Zhengyuan Wang, "Twitter Sentiment and IPO Performance: A Cross-Sectional Examination," *Journal of Portfolio Management*, vol. 42, no. 4 (Summer 2016): 129–135.

2 Steven L. Heston and Nitish Ranjan Sinha, "News vs. Sentiment: Predicting Stock Returns from News Stories," *Financial Analysts Journal*, vol. 73, no. 3 (Third Quarter 2017): 67–83. (https://www.cfapubs.org/doi/abs/10.2469/faj.v73.n3.3).

Robo-advice tends to start with an investor questionnaire, which may include many of the categories and subcategories shown in Exhibit 4. Exhibit 4 is a synthesis of questionnaires created by the researchers attributed in the source below. Once assets, liabilities, risk preferences, and target investment returns have been digitally entered by a client, the robo-adviser software produces recommendations, based on algorithmic rules and historical market data, that incorporate the client's stated investment parameters. According to research by Michael Tertilt and Peter Scholz, robo-advisers do not seem to incorporate the full range of available information into their recommendations;[3] further research will be necessary over time to see how this may affect performance and the evolution of digital advisory services. Nevertheless, current robo-advisory services include automated asset allocation, trade execution, portfolio optimization, tax-loss harvesting, and rebalancing for investor portfolios.

Exhibit 4 Categories and Subcategories for Investor Questionnaires

General Information	**Risk Tolerance**
Income	Age
Investment Amount	Association with Investing
Job Description	Association with Risk
Other	Choose Portfolio Risk Level
Source of Income	Comfort Investing in Stock
Spendings	Credit Based Investments
Time to Retirement	Dealing with Financial Decisions
Type of Account	Degree of Financial Risk Taken
Working Status	Education
Risk Capacity	Ever Interested in Risky Asset for Thrill
Dependence on Withdrawal of Investment Amount	Experience of Drop/Reaction on Drop/Max Drop before Selling
Income Prediction	Family and Household Status
Investment Amount/Savings Rate Ratio	Financial Knowledge
Investment Amount/Total Capital Ratio	Gender
Investment Horizon	Investment Experience
Liabilities	Investment Goal
Savings Rate	Investor Type/Self-Assessment Risk Tolerance
Total Capital	Preference Return vs. Risk

Source: Michael Tertilt and Peter Scholz, 2017 "To Advise, or Not to Advise—How Robo-Advisors Evaluate the Risk Preferences of Private Investors," working paper (13 June): Table 1: Categories and Subcategories for Questionnaires.

Although their analyses and recommendations can cover both active and passive management styles, most robo-advisers follow a passive investment approach. These robo-advisers typically have low fees and low account minimums, implementing their recommendations with low-cost, diversified index mutual funds or exchange-traded funds (ETFs). A diverse range of asset classes can be managed in this manner, including

3 Michael Tertilt and Peter Scholz, To Advise, or Not to Advise — How Robo-Advisors Evaluate the Risk Preferences of Private Investors (June 12, 2017). Available at SSRN: https://ssrn.com/abstract=2913178 or http://dx.doi.org/10.2139/ssrn.2913178

stocks, bonds, commodities, futures, and real estate. Because of their low-cost structure, robo-advisers can reach underserved populations, such as the mass affluent or mass market segments, which are less able to afford a traditional financial adviser.

Two types of wealth management services dominate the robo-advice sector: fully automated digital wealth managers and adviser-assisted digital wealth managers.

- **Fully Automated Digital Wealth Managers**

  The fully automated model does not rely on assistance from a human financial adviser. These services seek to offer a low-cost solution to investing and recommend an investment portfolio, which is often composed of ETFs. The service package may include direct deposits, periodic rebalancing, and dividend reinvestment options.

- **Adviser-Assisted Digital Wealth Managers**

  Adviser-assisted digital wealth managers provide automated investment services along with a virtual financial adviser, who is available to offer basic financial planning advice and periodic reviews by phone. Adviser-assisted digital wealth managers are capable of providing additional services that may involve a more holistic analysis of a client's assets and liabilities.

Wealthy and ultra-wealthy individuals typically have had access to human advisory teams, but there has been a gap in the availability and quality of advisers to serve investors with less wealth. The advent of robo-advisers offers a cost-effective and easily accessible form of financial guidance. In following a typically passive investment approach, research suggests that robo-advisers tend to offer fairly conservative advice.

However, critics of robo-advisers have wondered what would happen in a time of crisis, when people most often look to human expertise for guidance. It may not always be completely transparent why a robo-adviser chooses to make a recommendation or take a trading action that it did, unlike a human adviser who can provide his or her rationale. And finally, there may be trust issues in allowing computers to make these decisions, including worries of instances where robo-advisers might recommend inappropriate investments.

As the complexity and size of an investor's portfolio grows, robo-advisers may not be able to sufficiently address the particular preferences and needs of the investor. In the case of extremely affluent investors who may own a greater number of asset types—including alternative investments (e.g., venture capital, private equity, hedge funds, and real estate)—in addition to global stocks and bonds and have greater demands for customization, the need for a team of human advisers, each with particular areas of investment or wealth-management expertise, is likely to endure.

6.3 Risk Analysis

As mandated by regulators worldwide, the global investment industry has undertaken major steps in stress testing and risk assessment that involve the analysis of vast amounts of quantitative and qualitative risk data. Required data include information on the liquidity of the firm and its trading partners, balance sheet positions, credit exposures, risk-weighted assets, and risk parameters. Stress tests may also take qualitative information into consideration, such as capital planning procedures, expected business plan changes, business model sustainability, and operational risk.

There is increasing interest in monitoring risk in real time. To do so, relevant data must be taken by a firm, mapped to known risks, and identified as it moves within the firm. Data may be aggregated for reporting purposes or used as inputs to risk models. Big Data may provide insights into real-time and changing market circumstances to help identify weakening market conditions and adverse trends in advance, allowing managers to employ risk management techniques and hedging practices sooner to help preserve asset value. For example, evaluation of alternative data using ML techniques may help foreshadow declining company earnings and future stock performance. Furthermore, analysis of real-time market data and trading patterns may help analysts detect buying or selling pressure in the stock.

ML techniques may be used to help assess data quality. To help ensure accurate and reliable data that may originate from numerous alternative data sources, ML techniques can help validate data quality by identifying questionable data, potential errors, and data outliers before integration with traditional data for use in risk models and in risk management applications.

Portfolio risk management often makes use of scenario analysis—analyzing the likely performance of the portfolio and liquidation costs under a hypothetical stress scenario or the repeat of a historical stress event. For example, to understand the implications of holding or liquidating positions during adverse or extreme market periods, such as the financial crisis, fund managers may perform "what-if" scenario analysis and portfolio backtesting using point-in-time data to understand liquidation costs and portfolio consequences under differing market conditions. These backtesting simulations are often computationally intense and may be facilitated through the use of advanced AI-based techniques.

6.4 Algorithmic Trading

Algorithmic trading is the computerized buying and selling of financial instruments, in accordance with pre-specified rules and guidelines. Algorithmic trading is often used to execute large institutional orders, slicing orders into smaller pieces and executing across different exchanges and trading venues. Algorithmic trading provides investors with many benefits, including speed of execution, anonymity, and lower transaction costs. Over the course of a day, algorithms may continuously update and revise their execution strategy on the basis of changing prices, volumes, and market volatility. Algorithms may also determine the best way to price the order (e.g., limit or market order) and the most appropriate trading venue (e.g., exchange or dark pool) to route for execution.

High-frequency trading (HFT) is a form of algorithmic trading that makes use of vast quantities of granular financial data (tick data, for example) to automatically place trades when certain conditions are met. Trades are executed on ultra-high-speed, low-latency networks in fractions of a second. HFT algorithms decide what to buy or sell and where to execute on the basis of real-time prices and market conditions, seeking to earn a profit from intraday market mispricings.

Global financial markets have undergone substantial change as markets have fragmented into multiple trading destinations consisting of electronic exchanges, alternative trading systems, and so-called dark pools, and average trade sizes have fallen. In this environment, and with markets continuously reflecting real-time information, algorithmic trading has been viewed as an important tool.

DISTRIBUTED LEDGER TECHNOLOGY

7

Distributed ledger technology—technology based on a distributed ledger (defined below)—represents a fintech development that offers potential improvements in the area of financial record keeping. DLT networks are being considered as an efficient means to create, exchange, and track ownership of financial assets on a peer-to-peer basis. Potential benefits include greater accuracy, transparency, and security in record keeping; faster transfer of ownership; and peer-to-peer interactions. However, the technology is not fully secure, and breaches in privacy and data protection are possible. In addition, the processes underlying DLT generally require massive amounts of energy to verify transaction activity.

A **distributed ledger** is a type of database that may be shared among entities in a network. In a distributed ledger, entries are recorded, stored, and distributed across a network of participants so that each participant has a matching copy of the digital database. Basic elements of a DLT network include a digital ledger, a consensus mechanism used to confirm new entries, and a participant network.

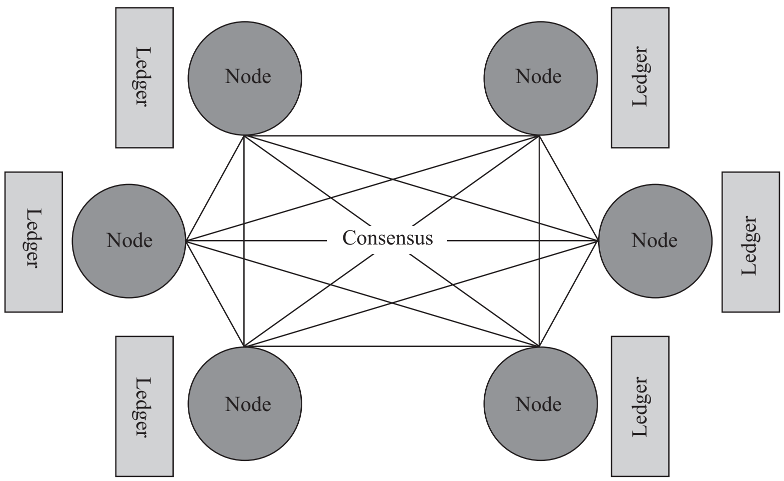
The consensus mechanism is the process by which the computer entities (or nodes) in a network agree on a common state of the ledger. Consensus generally involves two steps: transaction validation and agreement on ledger update by network parties. These features enable the creation of records that are, for the most part, considered immutable, or unchangeable, yet they are transparent and accessible to network participants on a near-real-time basis.

Features of DLT include the use of **cryptography**—an algorithmic process to encrypt data, making the data unusable if received by unauthorized parties—which enables a high level of network security and database integrity. For example, DLT uses cryptographic methods of proof to verify network participant identity and for data encryption.

DLT has the potential to accommodate "**smart contracts**," which are computer programs that self-execute on the basis of pre-specified terms and conditions agreed to by the parties to a contract. Examples of smart contract use are the automatic execution of contingent claims for derivatives and the instantaneous transfer of collateral in the event of default.

Exhibit 5 illustrates a distributed ledger network in which all nodes are connected to one another, each having a copy of the distributed ledger. The term "Consensus" is shown in the center of the network and represents the consensus mechanism in which the nodes agree on new transactions and ledger updates.

Exhibit 5 Distributed Ledger Network Setup



Source: https://blockgeeks.com/guides/what-is-hyperledger/.

Blockchain is a type of digital ledger in which information, such as changes in ownership, is recorded sequentially within blocks that are then linked or "chained" together and secured using cryptographic methods. Each block contains a grouping of transactions (or entries) and a secure link (known as a hash) to the previous block.

New transactions are inserted into the chain only after validation via a consensus mechanism in which authorized members agree on the transaction and the preceding order, or history, in which previous transactions have occurred.

The consensus mechanism used to verify a transaction includes a cryptographic problem that must be solved by some computers on the network (known as miners) each time a transaction takes place. The process to update the blockchain can require substantial amounts of computing power, making it very difficult and extremely expensive for an individual third party to manipulate historical data. To manipulate historical data, an individual or entity would have to control the majority of nodes in the network. The success of the network, therefore, relies on broad network participation.

Blockchain (Distributed Ledger) Network—How Do Transactions Get Added?

Outlined below are the steps involved in adding a transaction to a blockchain distributed ledger.

1 Transaction takes place between buyer and seller
2 Transaction is broadcast to the network of computers (nodes)
3 Nodes validate the transaction details and parties to the transaction
4 Once verified, the transaction is combined with other transactions to form a new block (of predetermined size) of data for the ledger
5 This block of data is then added or linked (using a cryptographic process) to the previous block(s) containing data
6 Transaction is considered complete and ledger has been updated

7.1 Permissioned and Permissionless Networks

DLT can take the form of permissionless or permissioned networks. **Permissionless networks** are open to any user who wishes to make a transaction, and all users within the network can see all transactions that exist on the blockchain. In a permissionless, or open, DLT system, network participants can perform all network functions.

The main benefit of a permissionless network is that it does not depend on a centralized authority to confirm or deny the validity of transactions, because this takes place through the consensus mechanism. This means no single point of failure exists, since all transactions are recorded on a single distributed database and every node stores a copy of the database. Once a transaction has been added to the blockchain, it cannot be changed, barring manipulation; the distributed ledger becomes a permanent and immutable record of all previous transactions. In a permissionless network, trust is not a requirement between transacting parties.

A well-known example of an open, permissionless network is **bitcoin**. Using blockchain technology, Bitcoin was created in 2009 to serve as the public ledger for all transactions occurring on its virtual currency. Since the introduction of bitcoin, many more cryptocurrencies, or digital currencies, which use permissionless DLT networks, have been created.

In **permissioned networks**, network members may be restricted from participating in certain network activities. Controls, or permissions, may be used to allow varying levels of access to the ledger, from adding transactions (e.g., a participant) to viewing transactions only (e.g., a regulator) to viewing selective details of the transactions but not the full record.

7.2 Applications of Distributed Ledger Technology to Investment Management

Potential applications of DLT to investment management include cryptocurrencies, tokenization, post-trade clearing and settlement, and compliance.

7.2.1 *Cryptocurrencies*

A **cryptocurrency**, also known as a digital currency, operates as electronic currency and allows near-real-time transactions between parties without the need for an intermediary, such as a bank. As electronic mediums of exchange, cryptocurrencies lack physical form and are issued privately by individuals, companies, and other organizations. Most issued cryptocurrencies utilize open DLT systems in which a decentralized distributed ledger is used to record and verify all digital currency transactions. Cryptocurrencies have not traditionally been government backed or regulated. Central banks around the world, however, are recognizing potential benefits and examining use cases for their own cryptocurrency versions.

Many cryptocurrencies have a self-imposed limit on the total amount of currency they may issue. Although such limits could help maintain their store of value, it is important to note that many cryptocurrencies have experienced high levels of price volatility. A lack of clear fundamentals underlying these currencies has contributed to their volatility.

Cryptocurrencies have proven to be an attractive means for companies looking to raise capital. An **initial coin offering** (ICO) is an unregulated process whereby companies sell their crypto tokens to investors in exchange for fiat money or for another agreed upon cryptocurrency. An ICO is typically structured to issue digital tokens to investors that can be used to purchase future products or services being developed by the issuer. ICOs provide an alternative to traditional, regulated capital-raising processes, such as initial public offerings (IPOs). Compared to the regulated IPO market, ICOs may have lower associated issuance costs and shorter capital raising time frames. However, most ICOs do not typically have attached voting rights. Regulation for ICOs is under consideration in a number of jurisdictions, and there have been numerous instances of investor loss resulting from fraudulent schemes.

7.2.2 *Tokenization*

Transactions involving physical assets, such as real estate, luxury goods, and commodities, often require substantial efforts in ownership verification and examination each time a transfer in ownership takes place. Through **tokenization**, the process of representing ownership rights to physical assets on a blockchain or distributed ledger, DLT has the potential to streamline this process by creating a single, digital record of ownership with which to verify ownership title and authenticity, including all historical activity. Real estate transactions that require ownership and identify verification may be one area to benefit from tokenization, because these transactions are typically labor intensive and costly, involving decentralized, paper-based records and multiple parties.

7.2.3 *Post-Trade Clearing and Settlement*

In the financial securities markets, post-trade processes to confirm, clear, and settle transactions are often complex and labor intensive, requiring multiple interactions between counterparties and financial intermediaries. DLT has the ability to streamline existing post-trade processes by providing near-real-time trade verification, reconciliation, and settlement, thereby reducing the complexity, time, and costs associated with processing transactions. A single distributed record of ownership between network peers would eliminate the need for independent and duplicative reconciliation efforts between parties and reduce the need for third-party facilitation. A shortened settlement time frame could lessen the time exposed to counterparty risk and associated collateral requirements while increasing the potential liquidity of assets and funds. Additionally, the use of automated contracts may also help to reduce post-trade time frames, lowering exposure to counterparty credit risk and trade fails.

7.2.4 *Compliance*

Regulators worldwide have imposed more stringent reporting requirements and demand greater transparency and access to data. To meet these requirements, many firms have added staff to their post-trade and compliance groups. But these functions remain predominantly manual. To comply with regulations, firms need to maintain and process large amounts of risk-related data. DLT may allow regulators and firms to maintain near-real-time review over transactions and other compliance-related processes. Improved post-trade reconciliation and automation through DLT could lead to more accurate record keeping and create operational efficiencies for a firm's compliance and regulatory reporting processes, while providing greater transparency and auditability for external authorities and regulators.

DLT-based compliance may better support shared information, communications, and transparency within and between firms, exchanges, custodians, and regulators. Closed or permissioned networks could offer advantages in security and privacy. These platforms could store highly sensitive information in a way that is secure but easily accessible to internal and external authorities. DLT could help uncover fraudulent activity and reduce compliance costs associated with know-your-customer and anti-money-laundering regulations, which entail verifying the identity of clients and business partners.

DLT Challenges

A number of challenges exist before DLT may be successfully adopted by the investment industry. These include the following:

- There is a lack of DLT network standardization, as well as difficulty integrating with legacy systems.
- DLT processing capabilities may not be financially competitive with existing solutions.
- Increasing the scale of DLT systems requires substantial (storage) resources.
- Immutability of transactions means accidental or "canceled" trades can be undone only by submitting an equal and offsetting trade.
- DLT requires huge amounts of computer power normally associated with high electricity usage.
- Regulatory approaches may differ by jurisdiction.

SUMMARY

- The term "fintech" refers to technological innovation in the design and delivery of financial services and products.

- Areas of fintech development include the analysis of large datasets, analytical techniques, automated trading, automated advice, and financial record keeping.
- Big Data is characterized by the three Vs—volume, velocity, and variety—and includes both traditional and non-traditional (or alternative) datasets.
- Among the main sources of alternative data are data generated by individuals, business processes, and sensors.
- Artificial intelligence computer systems are capable of performing tasks that traditionally required human intelligence at levels comparable (or superior) to those of human beings.
- Machine learning (ML) computer programs are able to "learn" how to complete tasks, improving their performance over time with experience. Main types of ML include supervised and unsupervised learning.
- Natural language processing is an application of text analytics that uses insight into the structure of human language to analyze and interpret text- and voice-based data.
- Robo-advisory services are providing automated advisory services to increasing numbers of retail investors. Services include asset allocation, portfolio optimization, trade execution, rebalancing, and tax strategies.
- Big Data and ML techniques may provide insights into real-time and changing market circumstances to help identify weakening or adverse trends in advance, allowing for improved risk management and investment decision making.
- Algorithmic traders use automated trading programs to determine when, where, and how to trade an order on the basis of pre-specified rules and market conditions. Benefits include speed of executions, lower trading costs, and anonymity.
- Blockchain and distributed ledger technology (DLT) may offer a new way to store, record, and track financial assets on a secure, distributed basis. Applications include cryptocurrencies and tokenization. Additionally, DLT may bring efficiencies to post-trade and compliance processes through automation, smart contracts, and identity verification.

PRACTICE PROBLEMS

1 A correct description of fintech is that it:
   A is driven by rapid growth in data and related technological advances.
   B increases the need for intermediaries.
   C is at its most advanced state using systems that follow specified rules and instructions.

2 A characteristic of Big Data is that:
   A one of its traditional sources is business processes.
   B it involves formats with diverse types of structures.
   C real-time communication of it is uncommon due to vast content.

3 In the use of machine learning (ML):
   A some techniques are termed “black box” due to data biases.
   B human judgment is not needed because algorithms continuously learn from data.
   C training data can be learned too precisely, resulting in inaccurate predictions when used with different datasets.

4 Text Analytics is appropriate for application to:
   A economic trend analysis.
   B large, structured datasets.
   C public but not private information.

5 In providing investment services, robo-advisers are *most likely* to:
   A rely on their cost effectiveness to pursue active strategies.
   B offer fairly conservative advice as easily accessible guidance.
   C be free from regulation when acting as fully-automated wealth managers.

6 Which of the following statements on fintech’s use of data as part of risk analysis is correct?
   A Stress testing requires precise inputs and excludes qualitative data.
   B Machine learning ensures that traditional and alternative data are fully segregated.
   C For real-time risk monitoring, data may be aggregated for reporting and used as model inputs.

7 A factor associated with the widespread adoption of algorithmic trading is increased:
   A market efficiency.
   B average trade sizes.
   C trading destinations.

8 A benefit of distributed ledger technology (DLT) favoring its use by the investment industry is its:
   A scalability of underlying systems.
   B ease of integration with existing systems.
   C streamlining of current post-trade processes.

9 What is a distributed ledger technology (DLT) application suited for physical assets?

A Tokenization

B Cryptocurrencies

C Permissioned networks

SOLUTIONS

1 A is correct. Drivers of fintech include extremely rapid growth in data (including their quantity, types, sources, and quality) and technological advances enabling the capture and extraction of information from it.

2 B is correct. Big Data is collected from many different sources and is in a variety of formats, including structured data (e.g., SQL tables or CSV files), semi-structured data (e.g., HTML code), and unstructured data (e.g., video messages).

3 C is correct. Overfitting occurs when the ML model learns the input and target dataset too precisely. In this case, the model has been "over trained" on the data and is treating noise in the data as true parameters. An ML model that has been overfitted is not able to accurately predict outcomes using a different dataset and may be too complex.

4 A is correct. Through the Text Analytics application of natural language processing (NLP), models using NLP analysis may incorporate non-traditional information to evaluate what people are saying—via their preferences, opinions, likes, or dislikes—in the attempt to identify trends and short-term indicators about a company, a stock, or an economic event that might have a bearing on future performance.

5 B is correct. Research suggests that robo-advisers tend to offer fairly conservative advice, providing a cost-effective and easily accessible form of financial guidance to underserved populations, such as the mass affluent and mass market segments.

6 C is correct. There is increasing interest in monitoring risk in real-time. To do so, relevant data must be taken by a firm, mapped to known risks, and identified while moving within the firm. Data may be aggregated for reporting purposes or used as inputs to risk models.

7 C is correct. Global financial markets have undergone substantial change as markets have fragmented into multiple trading destinations consisting of electronic exchanges, alternative trading systems, and so-called dark pools. In such an environment, when markets are continuously reflecting real-time information and continuously changing conditions, algorithmic trading has been viewed as an important tool.

8 C is correct. DLT has the potential to streamline the existing, often complex and labor intensive post-trade processes in securities markets by providing close to real-time trade verification, reconciliation, and settlement, thereby reducing related complexity, time, and costs.

9 A is correct. Through tokenization—the process of representing ownership rights to physical assets on a blockchain or distributed ledger—DLT has the potential to streamline this rights process by creating a single, digital record of ownership with which to verify ownership title and authenticity, including all historical activity.