Владимирский государственный университет

Т. И. КОЙКОВА

СПЕЦИАЛЬНОСТИ В СФЕРЕ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

MAJORS IN IT

Учебно-практическое пособие

Владимир 2025

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Владимирский государственный университет
имени Александра Григорьевича и Николая Григорьевича Столетовых»

Т. И. КОЙКОВА

СПЕЦИАЛЬНОСТИ В СФЕРЕ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

MAJORS IN IT

Учебно-практическое пособие

Электронное издание

Владимир 2025

ISBN 978-5-9984-1930-0

УДК 811.111
ББК 81.2. (Англ)

Рецензенты:
Кандидат педагогических наук, доцент
зав. кафедрой русского и иностранных языков юридического факультета
Владимирского юридического института ФСИН России
Е. Н. Романова

Кандидат педагогических наук, доцент
доцент кафедры второго иностранного языка и методики обучения
иностранным языкам Владимирского государственного университета
имени Александра Григорьевича и Николая Григорьевича Столетовых
М. В. Гайлит

Койкова, Т. И.

Специальности в сфере информационных технологий = Majors in IT [Электронный ресурс] : учеб.-практ. пособие / Т. И. Койкова ; Владим. гос. ун-т им. А. Г. и Н. Г. Столетовых. – Владимир : Изд-во ВлГУ, 2025. – 95 с. – ISBN 978-5-9984-1930-0. – Электрон. дан. (1,12 Мб). – 1 электрон. опт. диск (CD-ROM). – Систем. требования: Intel от 1,3 ГГц ; Windows XP/7/8/10 ; Adobe Reader ; дисковод CD-ROM. – Загл. с титул. экрана.

Включает в себя аутентичные тексты по темам: Data Science, Machine Learning, Artificial Intelligence, Cybersecurity, Neuro Networks и практические задания для работы в аудитории. Тексты раздела “Supplementary Reading” могут быть использованы для внеаудиторного чтения.

Предназначено для студентов и магистрантов направлений подготовки 09.03.01 «Информатика и вычислительная техника», 09.03.03 «Прикладная информатика», 09.03.02 «Информационные системы и технологии», 09.03.04 «Программная инженерия», 10.03.01 «Информационная безопасность», 10.05.04 «Информационно-аналитические системы».

Рекомендовано для формирования профессиональных компетенций в соответствии с ФГОС ВО.

ISBN 978-5-9984-1930-0

CONTENTS

ПРЕДИСЛОВИЕ

В условиях современности, когда динамично развивающиеся информационные технологии становятся частью нашей повседневной жизни, существует большая потребность в высококвалифицированных специалистах в области компьютерных коммуникаций. А знание английского языка для ИТ-профессионалов − это возможность оставаться в курсе последних новаций в области компьютерной техники, программного обеспечения, интернет-ресурсов, информационной безопасности, искусственного интеллекта и т. д.

Пособие предназначено для расширенного изучения английского языка в сфере информационных технологий для студентов и магистрантов, владеющих грамматикой и имеющих базовый запас английских лексических единиц. Цель пособия − развитие у учащихся навыков чтения специальной литературы на английском языке для извлечения необходимой информации, активизация лексического минимума, определенного для данного этапа обучения, что обеспечивает доступ к информации большинства ресурсов, документации и инструкций в области ИТ. Не последняя роль отводится формированию коммуникативных компетенций в рамках будущей специальности.

Пособие включает в себя девять уроков, некоторые из которых объединены общей темой (например, «Машинное обучение» и «Инженер по машинному обучению»). Комплекс упражнений, сопровождающий каждый текст, предполагает работу с отдельными абзацами с точки зрения их содержания и терминологии. Блок упражнений также содержит задания, направленные на выработку умений выстроить собственное высказывание.

Автор благодарит рецензентов *Елену Николаевну Романову*, кандидата педагогических наук, и *Марину Васильевну Гайлит*, кандидата педагогических наук, за ценные рекомендации по работе над пособием.

INTRODUCTION

The 21st century is called **the age of information technologies** due to the great influence of digital innovations, which have led to drastic changes in all aspects of the human life. The information age, also known as the Third Industrial Revolution, is a historical period that began in the mid-20th century. As any industrial revolution, this one is characterized by a rapid shift from traditional industries; this shift being to the knowledge-based economy centered on information technologies.

By the 1970s, with the development of the Internet by the United States Department of Defense and the subsequent adoption of personal computers a decade later, the Information, or Digital, Revolution was underway. There were some technological advances, which accelerated the transmission and processing of information. Among them was the development of fiber optic cables and faster microprocessors. The World Wide Web, which was initially used by companies as an electronic billboard for their products and services, transformed into an active consumer exchange for goods and information and became publicly accessible. Electronic mail, which permitted near-instant exchange of information, was widely adopted as the primary platform for workplace and personal communications. Today, we can message, call, learn about world events in real time, and access all forms of media in a matter of moments.

In the future, the rapid evolution of artificial intelligence, quantum computing, and biotechnology is anticipated. While IT will remain a dominant force, humanities will maintain a critical role by providing creativity, cultural preservation, and ethical perspectives.

Unit I
WHAT IS DATA SCIENCE?

Vocabulary
data science – аналитика данных
identify – определять / устанавливать тождество
opportunity – возможность
ultimately – в конечном итоге
predict – предсказывать
benefit – выгода, польза, преимущество
incorporate - объединять
data-savvy – разбирающийся в данных
delve – углублять
fraud (fraudulent) – мошенничество (мошеннический)
flawed – ошибочный
uptime – время безотказной работы
ROI (Return on Investment) – коэффициент окупаемости средств
to factor evidence – учитывать фактические данные

I. Data science is the field, where advanced analytic techniques and scientific principles are used for extracting valuable information from data and its further application for business decision-making, strategic planning and other areas. Nowadays, data science is considered to be increasingly important in businesses. The insights that data science generates help organizations make their operations more efficient, determine new business capabilities and improve marketing and sales programs, among other benefits. **Ultimately**, they can lead to competitive advantages over business rivals.

II. Data science **incorporates** various disciplines. Among them are data engineering, data preparation, data mining, **predictive analytics**, machine learning and data visualization, as well as statistics, mathematics and software programming. Skilled data scientists are mostly engaged in the above areas, although lower-level data analysts may also be involved. In addition, many institutions and enterprises now rely to some certain extent

on citizen data scientists[1], a group that can include business intelligence (BI) professionals, business analysts, **data-savvy** business users, data engineers and other workers who don't have a formal data science background.

III. This comprehensive guide to data science gives the explanation what it is, why it's important to organizations, how it works, the business **benefits** it offers and the challenges it creates. There is also an overview of data science applications, tools and techniques, in addition to the information on what data scientists do and the skills they need. Throughout the guide, there are **hyperlinks** to related TechTarget articles that **delve** more deeply into the topics considered here and offer insight and expert advice on data science initiatives.

IV. Data science plays an important role in virtually all aspects of business operations and strategies. For example, it provides information about customers that, in its turn, helps companies create effective marketing campaigns and targeted advertising, which is aimed at increasing product sales. It aids in managing financial risks, detecting **fraudulent transactions** and preventing equipment breakdowns at manufacturing plants and other industrial settings. It helps block cyber attacks and other security threats in IT systems.

V. From an operational standpoint, data science initiatives can optimize management of supply chains, product inventories, distribution networks and customer service. On a more fundamental level, they point the way to increased efficiency and reduced costs. Data science also enables companies to create business plans and strategies that are based on informed analysis of customer behavior, market trends and competition. Without it, businesses may miss **opportunities** and make **flawed decisions**.

VI. Data science is also vital in areas beyond regular business operations. In healthcare, its uses include diagnosis of medical conditions, image analysis, treatment planning and medical research. Academic institutions use data science to monitor student performance and improve their marketing to prospective students. Sports teams analyze player performance and plan game strategies via data science. Government agencies and public policy organizations are also big users.

VII. What are benefits of data science? In an October 2020 webinar organized by Harvard University's Institute for Applied Computational Science, Jessica Stauth, managing director for data science in the Fidelity Labs unit at Fidelity Investments, said there's "a very clear relationship" between data science work and business results. She cited potential business benefits that include higher **ROI**, sales growth, more efficient operations, faster time to market and increased customer engagement and satisfaction.

VIII. Generally speaking, one of data science's biggest benefits is to empower and facilitate better decision-making. Organizations that invest in it can **factor** quantifiable, data-based evidence into their business decisions. Ideally, such data-driven decisions will lead to stronger business performance, cost savings and smoother business processes and workflows.

IX. The specific business benefits of data science vary depending on the company and industry. In customer-facing organizations, for example, data science helps identify and refine target audiences. Marketing and sales departments can mine customer data to improve conversion rates and create personalized marketing campaigns and promotional offers that produce higher sales.

X. In other cases, the benefits include reduced **fraud**, more effective risk management, more profitable financial trading, increased manufacturing **uptime**, better supply chain performance, stronger cybersecurity protections and improved patient outcomes. Data science also enables real-time analysis of data as it's generated.

(https://www.analytixlabs.co.in/blog/what-is-data-science/)

Note:

The term **"Citizen Data Scientist"** (CDS for short) was for the first time used by the consulting company Gartner in 2018. It implies a specialist inside the company, who has the experience in some subject area and, at the same time, data analysis skills, which enables him to optimize some aspects in his subject area. At the companies where a technologist collaborates with a data scientist a CDS can do the job for two.

EXERCISES

I. Which of the paragraphs contain the following information?

1. There is a relationship between data science work and business results.
2. The role of data science in creating marketing campaigns and targeted advertising.
3. Data science helps business men become leaders in their competitive struggle.
4. Data science is a tool of countering cyberattacks.
5. A great number of scientists are involved in the field of Data science.
6. Insight and expert advice on data science initiatives can be found in the comprehensive guide to data science.
7. Business benefits of data science depend on the kind of the company.
8. Data science can be widely used in such areas as medicine and sport.
9. Data science gives an opportunity to make real-time analysis of data as it's generated.
10. Data science is efficient in making decisions in the process of managing a company.
11. Data science enables companies to optimize the schemes of supply.

II. Match the synonyms from two columns:

1. savvy	a) intellect
2. fraud	b) smart
3. chain	c) fake
4. intelligence	d) advantage
5. benefit	e) faulty
6. identify	f) integrate
7. predict	g) determine
8. incorporate	h) investigate
9. ultimately	i) discover
10. delve	j) forecast
11. detect	k)finally
12. flawed	l) order

III. Match the following terms with the appropriate definitions

1. citizen data scientist	a) is the process of identifying, assessing, and prioritizing potential risks, followed by implementing strategies to minimize or mitigate their impact.
2. fraud detection	b) a settlement which often arises from time constraints during crises, which can prevent thorough analysis and lead to snap judgments.
3. data mining	c) a set of methodologies that enables systems to automatically learn and improve from various analyses and outputs without being explicitly programmed.
4. risk management	d) is an individual who does some data science work for an organization but doesn't hold the title of data scientist or have a formal background in advanced analytics, statistics or related disciplines.
5. supply chain	e) a set of business intelligence (BI) technologies that uncovers relationships and patterns within large volumes of data that can be used to predict behavior and events.

6. hyperlink	f) the unauthorized use of an individual's accounts or payment information. It can result in the victim's loss of funds, personal property, or personal information.
7. machine learning	g) is a computational process for discovering patterns, correlations, and anomalies within large datasets. It applies various statistical analysis and machine learning (ML) techniques to extract meaningful information and insights from data.
8. predictive analytics	h) is the sequence of processes involved in producing and distributing a commodity.
9. fraudulent transaction	i) is a word, phrase, or image that is linked to another document or website.
10. flawed decision	j) is the process of recognizing unauthorized activities where money or property is obtained through false pretenses, such as phishing, stolen credit cards, and identity theft.
11. uptime	k) the ratio between net income (over a period) and investment (costs resulting from an investment of some resources at a point in time).

12. ROI	l) the ratio of the total time during which a machinery or equipment is operational or the production time to the total available time

IV. Formulate the following word combinations and statements from the text in another way.

1. Data science is considered to be increasingly important in businesses.
2. Ultimately, they can lead to competitive advantages over business rivals.
3. data-savvy_business users
4. Skilled data scientists are mostly engaged in the above areas although lower-level data analysts may also be involved.
5. Effective marketing campaigns and targeted advertising are aimed at increasing product sales
6. It helps block cyber attacks and other security threats in IT systems.
7. Businesses may miss opportunities and make flawed decisions.
8. There's "a very clear relationship" between data science work and business results.
9. Organizations that invest in it can factor quantifiable, data-based evidence into their business decisions.
10. Data science also enables real-time analysis of data as it's generated.
11. Organizations can factor quantifiable, data-based evidence.
12. customer facing organization

V. Answer the following questions:

1. What are the main tools of data science used for receiving necessary information from data?
2. What is the contribution of data science in business?
3. What disciplines does data science involve?
4. What is the main idea of the present comprehensive guide?
5. What differs a citizen data scientist from other specialists?

6. How does data science facilitate marketing campaigns?
7. How can data science be used in the fields of healthcare and education?
8. What is one of the biggest benefits of data science?
9. What is the dependence of the data science benefits on the company profile?

VI. Speak on the following topics:

1. The definition of data science and the disciplines incorporated in it.
2. The benefits provided by data science.

Unit II
DATA SCIENTIST

Vocabulary

in a timely manner – своевременно
SQL (Structured Query Language) – язык структурированных запросов
dashboard – приборная панель
AI system (artificial intelligence system) – система искусственного интеллекта
a must – требование
expertise – компетентность
big data - супер массив данных
NLP- обработка текста на естественном языке (с лингвистической обработкой)
JSON — текстовый формат обмена данными, основанный на JavaScript.
disparate – различный
metrics –количественные показатели
ad hoc –специальный
insights – результаты аналитической обработки
steer – управлять

promote – содействовать
autonomous – автономный
trait – черта (характера)
glean – собирать (воедино)
maintain – подчеркивать
oversee – осуществлять контроль
machine learning framework – платформа машинного обучения
conversational AI system – диалоговая система искусственного интеллекта

I. A data scientist is a professional in analytics. He (she) collects, analyzes and interprets data, which is vital in making decision policy of any organization. The data scientist activities include elements of several traditional and technical jobs, including mathematician, scientist, statistician and computer programmer. Data scientists must be able to complete a great number of complex planning, modeling and analytical tasks **in a timely manner**. Given that, the job requires knowledge of various data science tools and libraries; big data platforms, such as Spark, Kafka, Hadoop and Hive; and programming languages that include Python, R, Julia, Scala and **SQL**.
II. In the process of their activities, data scientists also use advanced analytics techniques, such as machine learning and predictive modeling, along with the application of scientific principles. The basic responsibilities of a data scientist include the following activities: gathering and preparing relevant data to use in analytics applications; using various types of analytics tools to detect patterns, trends and relationships in data sets; developing statistical and predictive models to run against the data sets; and creating data visualizations, **dashboards** and reports to communicate their findings.
III. Data scientists often have to work with large amounts of data when they deal with developing and testing hypotheses, making inferences and analyzing things such as market trends, financial risks, cybersecurity threats, stock trades, equipment maintenance needs and medical conditions.
IV. Technical skills required for the job include data mining, predictive modeling, machine learning and deep learning, as well as upfront data processing and data preparation. The ability to work with a combination of structured,

semistructured and unstructured data is often a requirement, too, especially in big data environments that contain different types of data. Experience with statistical research and analytics techniques such as classification, clustering, regression and segmentation -- is also **a must**. In some cases, **expertise** in natural language processing (NLP) is another prerequisite.

Examples of necessary skills listed in job postings include the following:

1) expertise in all phases of data science, from initial data discovery through data cleansing and model selection, validation and deployment;
2) knowledge and understanding of common data warehouse and data lake structures;
3) experience with using statistical approaches to solve analytics problems;
4) proficiency in popular machine learning frameworks;
5) familiarity with common data science and machine learning techniques, such as decision trees, K-nearest neighbors, naive Bayes classifiers, random forests and support vector machines;
6) experience with techniques for both qualitative and quantitative analysis;
7) the ability to identify new opportunities to apply machine learning and data mining tools to business processes to improve their efficiency and effectiveness;
8) experience with public cloud platforms and services;
9) familiarity with a wide variety of data sources, including databases and big data platforms, as well as public or private APIs and standard data formats, like **JSON**, YAML and XML;
10) the ability to aggregate data from **disparate** sources and prepare it for analysis;
11) experience with data visualization tools, such as Tableau and Power BI;
12) the ability to design and implement reporting dashboards that can track key business **metrics** and provide actionable insights;
13) the ability to do **ad hoc** analysis and present the results in a clear manner.

V. Data scientists working for businesses typically mine data that could be used to predict customer behavior. They identify new revenue opportunities, detect fraudulent transactions and meet other business needs. They might also be asked to explore data without being given a specific business problem to solve. In that scenario, data scientists need to understand both

the data and the business well enough to formulate questions, do the analysis work and deliver **insights** to business executives on possible changes to business operations, products or services. They also need leadership capabilities and business savvy to help **steer** data-driven decision-making processes in an organization.

VI. Data scientists are often involved in valuable analytics work for healthcare providers, academic institutions, government agencies and other types of organizations. In many organizations, data scientists are also responsible for helping to define and **promote** best practices for data collection, preparation and analysis. In addition, some data scientists develop AI technologies for use internally or by customers -- for example, conversational AI systems, AI-driven robots and other **autonomous** machines, including key components in self-driving cars.

VII. Highly qualified data scientists possess such personality **traits** as the combination of their intellect with skepticism and intuition. They are tireless problem-solvers driven to find a needle in a haystack. They should be creative so that to guide further investigation with the goal of uncovering new information. Advanced data scientists should be good storytellers who know how to present data insights and who can communicate with people at all levels of an organization.

VIII. The term “data scientist” was first used as a job title in 2008, simultaneously at Facebook and LinkedIn. The demand for data science skills has grown significantly over the years, as companies are interested in **gleaning** useful information from great volumes of big data and taking advantage of artificial intelligence (AI) and machine learning technologies to enable new types of analytics applications.

IX. The role of data scientist is often confused with that of data analyst. There is really an overlap in many of the job responsibilities and required skills, but there are also some differences between data scientists and data analysts. The duties of a data analyst can vary depending on the company. In general, though, they don't have the full level of technical skills that data scientists need, and they might also be less experienced. They still collect, process and analyze data, as well as creating visualizations and dashboards

to report findings; some data analysts also design and **maintain** the databases and other data stores used in analytics applications.
However, data analysts often support the work of data scientists and are **overseen** by them in analytics initiatives.
(https://www.techtarget.com/searchenterpriseai/definition/data-cientist)

Notes:

1. **API (application programming interface)** is a connection between computers or between computer programs. It is a type of software interface, offering a service to other pieces of software. A document or standard that describes how to build such a connection or interface is called an API specification. A computer system that meets this standard is said to implement or expose an API. The term API may refer either to the specification or to the implementation.
2. **JSON (JavaScript Object Notation)** is an open standard file format and data interchange format that uses human-readable text to store and transmit data objects consisting of name–value pairs and arrays (or other serializable values).
3. **YAML** is a widely used format for writing configuration files for different DevOps tools, programs, and applications because of its human-readable and intuitive syntax.
4. **XML (Extensible Markup Language)** is a set of rules for encoding scripts (documents) in a way that is both understandable and machine-decipherable.

EXERCISES

I. Which of the paragraphs contain the following information?

1. Some data scientists develop AI technologies
2. The areas, which demand from the data scientist to work with large amounts of data.
3. Data scientist versus data analyst
4. Data scientists must have the knowledge of programming languages.

5. Characteristics of an effective data scientist
6. The activities, which a data scientist is responsible for.
7. The appearance of the term “a data scientist” as a job title.
8. The appearance of the term “a data scientist” as a job title.
9. On some occasions, data scientists need to understand both the data and the business.

II. Match the synonyms from two columns:

1. must	a) support
2. expertise	b) demand
3. disparate	c) specific
4. metrics	d) feature
5. ad hoc	e) various
6. steer	f) gather
7. promote	g) monitor
8. autonomous	h) competence
9. trait	i) encourage
10. glean	j) manage
11. maintain	k) independent
12. oversee	l) figures

III. Match the following terms with the appropriate definitions

1. A data scientist	a) a subfield of Artificial Intelligence (AI) and computer science that uses data and algorithms to mimic human learning processes and gradually improve accuracy.
2. A data lake	b) a centralized repository that stores structured data (database tables, Excel sheets) and semi-structured data (XML files, webpages)

	for the purposes of reporting and analysis.
3. Deep learning	c) the ability of a computer program to understand human language as it's spoken and written -- referred to as natural language.
4. A data warehouse	d) a combination of structured, semi-structured and unstructured data collected by organizations that can be mined for information and used in machine learning projects, predictive modeling and other advanced analytics applications.
5. Big data	e) a machine-based system that can operate autonomously and adapt after deployment, generating outputs like predictions or decisions.
6. Machine learning	f) is a type of machine learning that uses artificial neural networks to learn from data.
7. Natural language processing (NLP)	g) refers to the idea that having as much relevant information it helps a buyer make a better-informed decision on a property.

8. An AI system	h) refers to large, diverse sets of information that grow at ever-increasing rates. The term encompasses the volume of information, the velocity or speed at which it is created and collected, and the variety or scope of the data points being covered.
9. Upfront data	i) a detective of the digital age who uses a mix of statistics, computer science, and domain expertise to find hidden patterns and insights within massive amounts of data

IV. Formulate the following word combinations and statements from the text in another way.

1. Data scientists must be able to complete a great number of complex planning, modeling and analytical tasks in a timely manner.
2. Given that, the job requires knowledge of various data science tools and libraries
3. The basic responsibilities of a data scientist include the following activities: gathering and preparing relevant data …
4. In some cases, expertise in natural language processing (NLP) is another prerequisite.
5. initial data discovery
6. data mining tools
7. the ability to aggregate data from disparate sources and prepare it for analysis
8. the ability to do **ad hoc** analysis and present the results in a clear manner
9. deliver insights to business executives on possible changes to business operations
10. They are tireless problem-solvers driven to find a needle in a haystack.

11. Companies are interested in gleaning useful information from great volumes of big data and taking advantage of artificial intelligence (AI) and machine learning technologies to enable new types of analytics applications
12. They don't have the full level of technical skills that data scientists need

V. Answer the following questions:

1. What knowledge does the data scientist's job require?
2. What are the basic activities, for which data scientists are responsible?
3. When do data scientists have to work with large amounts of data?
4. What technical skills are required from data scientists?
5. Why do the data scientists working for businesses need to be good at both data and business?
6. Why is it desirable for the data scientists working in business possess leadership capabilities?
7. Why do educational and governmental institutions involve data scientists?
8. What personality traits do highly qualified data scientists have? Why should they be creative and sociable?
9. What encouraged the occurrence of data scientist as a job title?
10. What is the difference between data scientist and data analyst?

VI. Speak on the following topics.

1. The role of a data scientist in business.
2. Necessary technical skills of a data scientist.
3. Data scientist versus data analyst.

Unit III
KEY ASPECTS OF A DATA SCIENTIST'S JOB

Vocabulary
cleansing – очищение (purification)
validate – проверять достоверность (verify)
data set – пакет информации (package)
prep phase – подготовительный этап (stage)
rival – конкурент (competitor)
anomaly - отклонение (deviation)
to uncover patterns – выявить закономерность
outcome – (итоговый) результат (result)
predict – прогнозировать (forecast)
data silo – обособленная база данных
rigorous – тщательный (thorough)
inconsistent data – несогласованные данные
skew - искажать
upfront – заранее

I. The first step in data science applications is to collect and prepare the data that will be analyzed. Data preparation is the process of gathering, **cleansing**, organizing, transforming and **validating data sets** for analysis. Data scientists often work together with data engineers during the data **prep phase**.

II. Analyzing data to identify trends, correlations, **anomalies** and other useful information is the main purpose of data science initiatives. Overall, the analytics work done by data scientists is aimed at improving business performance and helping organizations gain a competitive advantage over business **rival**s.

III. As part of data analytics efforts, this involves working **to uncover patterns** and relationships in large data sets. Data mining typically is done by applying advanced algorithms to the data that is being analyzed. Data scientists then use the results generated by the algorithms to create analytical models.

IV. Increasingly, data mining and analytics are driven by machine learning, in which algorithms are built to learn about data sets and then find the desired information in them. Data scientists are responsible for training and overseeing machine learning algorithms as needed. Deep learning is a more advanced form that uses artificial neural networks.

V. Data scientists commonly also must be able to create predictive models of different business scenarios to analyze potential **outcomes** and behavior. For example, models can be built to **predict** how different customers will likely respond to marketing offers or to assess the possible indicators of diseases.

VI. Data science work also involves the use of statistical analysis techniques to analyze data sets. Statistical analysis is a core facet of what data scientists do to explore data and find underlying trends and patterns for analysis and interpretation.

VII. The findings of data science applications are usually organized into charts or other types of data visualizations so business executives and workers can easily understand them. In many cases, data scientists combine multiple visualizations into reports, interactive dashboards or detailed data stories.

VIII. Although they have what is considered to be one of the best jobs available, data science work is generally complex because of its advanced nature and the large amount of data that often must be analyzed. Also, because data scientists aren't always given specific analytics questions to answer or directions on how to focus their research, it sometimes can be hard to ensure that what they do meets business needs.

IX. Gathering relevant data for analytics applications can be difficult, too, especially in organizations with **data silos** that are isolated from other IT systems. Incorrect or inconsistent data can erroneously **skew** the results of analytics models; to avoid that, **rigorous** data profiling and cleansing is required **upfront** to identify and fix data quality issues. Overall, data preparation is time-consuming: A common maxim is that data scientists spend 80% of their time finding and preparing data and only 20% analyzing it.

X. Identifying and addressing biases in data science applications is another big challenge, both in the data being analyzed and in algorithms and

analytical models. Maintaining models and ensuring that they're updated when data sets or business requirements change can also be problematic. And analytics workloads might be hard to handle if companies don't invest in a full data science team.

Most data science jobs require at bare minimum a bachelor's degree in a technical field. More commonly, though, data scientists have an advanced degree in statistics, data science, computer science or mathematics.

(https://www.simplilearn.com/data-scientist-job-description-article)

Note:

to address a bias – устранить отклонение (смещение)

EXERCISES

I. Which of the paragraphs contain the following information?

1. The stage of a data scientist's job when artificial neural networks are used
2. The challenges, which data scientists face.
3. The stage which is considered to be the first stage of a data scientist's work
4. Educational background of a data scientist.
5. The aspect of a data scientist's job aimed at gaining a competitive advantage over business competitors
6. Relation between preparatory and analysis stages.
7. The phase of the job when a data scientist applies advanced algorithms to the analyzed data
8. The models, which are created for getting the knowledge about possible reaction of the customers to different offers.
9. The aspect, which is known to be the major one in to exploring data and finding the main trends and patterns for analysis.
10. The forms, which are used by the data scientists for presenting their findings.

II. Match the synonyms from two columns:

1. cleansing	a) beforehand
2. validate	b) forecast
3. data set	c) result
4. prep phase	d) competitor
5. rival	e) thorough
6. anomaly	f) purification
7. outcome	g) stage
8. predict	h) verify
9. rigorous	i) deviation
10. upfront	j) package

III. Match the following terms with the appropriate definitions

1. A dataset	a) a well-defined sequence of steps focused on solving a particular problem.
2. Anomaly	b) a mathematical process used to forecast future events or outcomes by analyzing patterns in a given set of input data.
3. Rival	c) a deviation from the common or expected; something out of the ordinary or simply does not fit in.
4. An algorithm	d) not merely a tool to render data into colorful graphs or charts but an artistic science, where data dons the cloak of visual elements such as shapes, colors, and patterns, revealing patterns, trends, and insights that might otherwise remain obscured.

5. Predictive modeling	e) a collection of data typically organized in tables, arrays or specific formats—such as CSV or JSON—for easy retrieval and analysis, essential for data analysis, machine learning (ML), artificial intelligence (AI) and other applications that require reliable, accessible data.
6. Visualization	f) a statistical phrase that represents a systematic variation from the real value. It is the difference between the desired value and the actual value of the parameter.
7. Data silo	g) a person or group that tries to defeat or be more successful than another person or group
8. Bias	h) a pocket of information stored in different information systems or subsystems that don't connect with one another

IV. Formulate the following word combinations and statements from the text in another way.

1. Overall, the analytics work done by data scientists is aimed at improving business performance and helping organizations gain a competitive advantage over business rivals.
2. Increasingly, data mining and analytics are driven by machine learning.

3. … models can be built to predict how different customers will likely respond to marketing offers or to assess the possible indicators of diseases.
4. Statistical analysis is a core facet of what data scientists do to explore data and find underlying trends and patterns for analysis and interpretation.
5. Incorrect or inconsistent data can erroneously skew the results of analytics models.
6. Rigorous data profiling and cleansing is required upfront to identify and fix data quality issues.
7. to identify and address biases in data science applications
8. Maintaining models and ensuring that they're updated when data sets or business requirements change can also be problematic.

V. Answer the following questions.

1. What activities are included in the preparation stage?
2. What is the main purpose of the data scientists' initiatives?
3. What do data scientists use for creating analytical models?
4. What are data scientists responsible for when speaking about machine learning?
5. What is the difference between machine learning and deep leaning?
6. What are predictive models created for?
7 What are the main forms of visualization of the data scientists' findings?
8. What difficulties do data scientists face?
9. What are the reasons of the challenges which data scientist have to solve?

VI. Speak on the following topics:

1. Key aspects of data scientist's job.
2. Challenges and complications experienced by data scientists.

Unit IV
WHAT IS DATA PROFILING?

Vocabulary
consistency – систематичность
timeliness – актуальность
compliance –соответствие
enhance – улучшать, совершенствовать
facilitate – способствовать
ensure– обеспечивать
adhere to smth. – соблюдать что-либо; придерживаться чего-либо
discrepancy – несоответствие, расхождение
data consistency – согласованность данных
proactively – (здесь) заранее
insights into - глубокое понимание (чего-либо)
open-source tool – инструментальное средство с открытым исходным кодом
robust – крепкий, устойчивый
duplicate – повторяющийся
outlier – резко отличающееся значение
constraint – ограничение
scalability –масштабирование
risk mitigation – снижение рисков
benchmark – ориентир
seamless – бесперебойный
emerge – возникать

I. Data profiling examines and assesses the quality, structure, and content of a dataset. This process plays a crucial role in data management by ensuring accuracy, **consistency**, and **timeliness**. Data profiling offers numerous benefits, such as improving decision-making, supporting **compliance** efforts, and reducing operational costs. Various industries, including finance, healthcare, and retail, utilize data profiling **to enhance** customer service, **facilitate** mergers and acquisitions, and improve data security. By

understanding the data's content and quality, organizations can build new products, solutions, or *data pipelines* more effectively.

II. Data profiling involves examining and assessing the quality, structure, and content of a dataset. This systematic analysis uncovers patterns, relationships, and anomalies within the data. Organizations use data profiling to verify data characteristics and **ensure compliance** with business rules and statistical standards. The process helps identify data quality issues, such as inconsistencies, null values, and incoherent schema designs.

III. Data profiling plays a crucial role in data management. It ensures high-quality and reliable data, which is essential for accurate decision-making. By identifying and fixing data quality problems early, organizations can avoid costly errors during data analysis. Data profiling supports compliance efforts and reduces operational costs. It also informs the creation of data quality rules that monitor and cleanse data continuously.

Key Concepts

IV. Data quality refers to the accuracy, completeness, and reliability of data. High-quality data meets the specific needs of an organization and **adheres to** defined standards. Data profiling helps ensure data quality by identifying errors, inconsistencies, and gaps in the dataset. This process enables organizations to maintain accurate and trustworthy data for analysis and reporting.

V. Data consistency ensures that data remains uniform across different datasets and systems. Inconsistent data can lead to incorrect analysis and poor decision-making. Data profiling helps detect and resolve **discrepancies** in data values and formats. By maintaining data consistency, organizations can improve the reliability of their data and enhance overall *data integrity*.

VI. Data completeness measures the extent to which all required data is present in a dataset. Incomplete data can hinder analysis and lead to inaccurate conclusions. Data profiling identifies missing values and incomplete records, allowing organizations to address these issues **proactively**. Ensuring data completeness enhances the overall quality and usability of the data.

Processes Involved in Data Profiling

VII. Data profiling begins with identifying sources of data. Organizations collect data from various sources. These sources include databases, spreadsheets, and cloud storage. External sources such as social media, public records, and third-party vendors also provide valuable data. Each source offers unique **insights** and contributes to a comprehensive dataset.

VIII. Organizations use different methods to collect data. Manual entry involves human input, which can introduce errors. Automated systems reduce errors by capturing data directly from digital sources. Web scraping extracts data from websites. APIs allow seamless data transfer between systems. Each method has advantages and limitations, influencing the quality and completeness of the collected data.

IX. Data analysis involves several techniques. Statistical analysis examines data distributions and identifies patterns. Data visualization presents data graphically, making trends and outliers easily identifiable. Machine learning algorithms detect complex patterns and predict future outcomes. Each technique provides unique insights, enhancing the understanding of the dataset.

X. Various tools assist in data analysis. **Open-source tools** like R and Python offer flexibility and customization. Commercial tools such as SAS and Tableau provide **robust** features and support. Data profiling tools like Talend and Informatica specialize in examining data quality and structure. The choice of tool depends on the specific needs and resources of the organization.

XI. Creating data profiles involves summarizing key characteristics of the dataset. Profiles include information on data types, distributions, and relationships. Data profiling tools generate these profiles automatically. Analysts review the profiles to identify anomalies and areas for improvement. This step ensures that the data meets quality standards before further analysis. Interpreting data profiles requires expertise. Analysts examine the profiles to understand the dataset's strengths and weaknesses. Patterns and anomalies provide **insights into** data quality issues. Analysts recommend

actions to address identified problems. Effective interpretation of data profiles leads to better decision-making and improved data management practices.

Types of Data Profiling

XII. Structure profiling examines the format and organization of data within

a dataset. This type of profiling focuses on metadata, such as data types, lengths, and **constraints**. Structure profiling helps identify schema inconsistencies and structural anomalies. For example, analysts may discover that a column intended for numerical data contains text values. This discovery prompts necessary corrections to maintain data integrity.

Content profiling analyzes the actual data values within a dataset. This profiling type assesses data quality by examining patterns, distributions, and anomalies. Content profiling helps identify issues like **duplicate records**, missing values, and **outliers**. For instance, content profiling might reveal that customer names contain special characters or inconsistent formats. Addressing these issues ensures accurate and reliable data for analysis.

Relationship profiling explores the connections between different datasets or tables. This type of profiling identifies relationships such as primary keys, foreign keys, and dependencies. Relationship profiling helps ensure data consistency across related datasets. For example, relationship profiling might uncover that a foreign key in one table does not match any primary key in another table. The resolution of these discrepancies improves data integrity and coherence.

Techniques and Tools for Data Profiling

XIII. Statistical analysis plays a crucial role in data profiling. Analysts use statistical methods to examine data distributions and identify patterns. These methods include measures of central tendency, variability, and correlation. By applying statistical techniques, data analysts can detect anomalies and outliers that may indicate data quality issues. Statistical analysis provides a quantitative foundation for understanding the dataset's characteristics.

Pattern recognition involves identifying **recurring** sequences or trends within the data. This technique helps uncover hidden relationships and dependencies. Analysts use pattern recognition to detect inconsistencies and irregularities in data values. For example, recognizing a pattern in customer purchase behavior can reveal insights into consumer preferences. Pattern recognition enhances the ability to predict future trends and make informed decisions based on historical data.

Open-source tools offer flexibility and customization for data profiling tasks. R and Python are popular choices due to their extensive libraries and community support. Tools like Pandas and Dplyr provide powerful data manipulation capabilities. Open-source tools allow organizations to tailor their data profiling processes to specific needs. The cost-effectiveness of open-source solutions makes them accessible to organizations of all sizes.

Commercial tools provide robust features and support for data profiling. Tools such as SaS and Tableau offer comprehensive data analysis and visualization capabilities. These tools often come with built-in functionalities for data quality assessment and reporting. Commercial tools streamline the data profiling process by automating many tasks. Organizations benefit from the reliability and **scalability** of these professional-grade solutions.

Benefits of Data Profiling

XIV. Data profiling enhances the accuracy of datasets. Organizations can identify and correct errors in data entries. This process ensures that the information used for analysis is precise. Accurate data supports reliable decision-making and operational efficiency.

Consistency in data is crucial for maintaining integrity across systems. Data profiling helps detect discrepancies in data formats and values. By resolving these inconsistencies, organizations can ensure uniformity in their datasets. Consistent data improves the reliability of reports and analyses.

Data profiling provides valuable insights into the dataset's characteristics. These insights help organizations understand patterns and trends within their data. Companies can leverage this information to make informed decisions. Data-driven insights lead to better strategies and improved business outcomes.

Risk mitigation is a critical benefit of data profiling. By identifying data quality issues early, organizations can prevent costly errors. Data profiling supports compliance efforts by ensuring that data adheres to regulatory standards. This proactive approach reduces the risk of non-compliance and associated penalties.

Challenges in Data Profiling

XV. Organizations often collect vast amounts of data. Managing these large datasets poses significant challenges. Analysts must process and analyze this data efficiently. Specialized tools become essential for handling such volumes. Without these tools, the task becomes overwhelming. Large datasets require robust infrastructure to store and manage data effectively.

Unstructured data lacks a predefined format. This type of data includes text, images, and videos. Analyzing unstructured data presents unique challenges. Traditional data profiling techniques may not apply. Advanced methods and tools are necessary to extract meaningful insights. Unstructured data often requires more time and resources for processing.

Data profiling demands considerable time and financial investment. Organizations must allocate resources for data collection and analysis. Hiring skilled professionals adds to the cost. The process can become expensive, especially for large-scale projects. Efficient planning and budgeting are crucial to manage these expenses.

Data profiling requires expertise in data analysis and management. Skilled personnel are essential for accurate and reliable results. Finding and retaining such talent can be challenging. Training existing staff may also be necessary. The need for specialized skills increases the overall cost and complexity of the process.

Best Practices for Effective Data Profiling

XVI. Organizations must define clear goals before starting data profiling. Goals provide direction and focus for the data profiling process. Clear objectives help identify what needs to be achieved. Organizations can prioritize tasks and allocate resources effectively. Setting **benchmarks** is essential for measuring progress. Benchmarks serve as reference points for evaluating data quality. Organizations can compare current data against these standards. This comparison helps identify areas that need improvement.

Benchmarks ensure that data profiling efforts align with organizational goals.

Selecting the right tools is crucial for effective data profiling. Organizations should consider several criteria when choosing tools. These criteria include functionality, ease of use, and cost. Tools must meet the specific needs of the organization. The right tools enhance the efficiency and accuracy of data profiling.

Integration with existing systems is vital for **seamless** data profiling. Tools should work well with the organization's current infrastructure. This **compatibility** ensures smooth data flow between systems. Integrated tools reduce the risk of data loss or corruption. Proper integration enhances the overall effectiveness of data profiling.

Continuous improvement is key to maintaining data quality. Organizations should regularly monitor data profiling results. This monitoring helps identify new issues and track progress. Continuous improvement ensures that data remains accurate and reliable. Regular updates to data profiles keep the information relevant and useful.

Adapting to changes is necessary for effective data profiling. Data environments constantly evolve, requiring flexibility. Organizations should adjust their data profiling processes as needed. This adaptability helps address **emerging** data quality issues. Staying responsive to changes ensures ongoing data integrity and reliability.

(https://www.ibm.com/think/topics/data-profiling)

EXERCISES

I. Which of the paragraphs contain the following information?

1. Methods of data collection reducing errors.
2. Some information about the creation of data profiles.
3. The need for specialized skills increases the overall cost and complexity of data profiling.

4. The technics used for detecting inconsistencies and irregularities in data values
5. It is very important to adapt to constantly evolving data environments.
6. Techniques for analyzing data
7. The means providing provide powerful data manipulation capabilities
8. The tools chosen for effective data profiling must meet some definite criteria.
9. The use of data profiling by organizations.
10. The concept of data profiling, which provides the compliance in data values and formats
11. Time and cost factors of data profiling
12. Regular monitoring and updating provide effective data profiling.
13. The importance of data profiling for making accurate decisions
14. The necessity to have special tools for handling large datasets.
15. The process, which enables organizations to maintain accurate and trustworthy data
16. Incomplete data can impede analysis and lead to inexact conclusions.
17. Tools used in the process of data analysis
18. The type of profiling, which identifies relationships such as primary keys, etc.
19. Data profiling helps detect discrepancies in data formats and values.
20. The dependence of decision-making on the effective interpretation of data profiles
21. The information about the types of data sources
22. The role of data-driven insights in enhancing decision-making.

II. Match the synonyms from two columns:

1. consistency	a) scaling duplication
2. timeliness	b) analytical data
3. compliance	c) inconsistency
4. enhance	d) follow smth.
5. facilitate	e) arise
6. ensure	f) improve

7. adhere (to)	g) encourage/promote
8. discrepancy	h) restriction
9. proactively	i) recurring
10. insights	j) actuality
11. robust	k) smooth
12. duplicate	l) reference point
13.constraint	m) regularity
14.scalability	n) compatibility
15.benchmark	o) in advance
16.seamless	p) durable
17.emerge	q) provide

III. Match the following terms with the appropriate definitions

1. A data science pipeline	a) series of interconnected steps and processes that transform raw data into valuable insights. It is an end-to-end framework, that takes data through various stages of processing, leading to actionable outcomes.
2. Data quality	b) refers to the reliability, accuracy, consistency and validity of your data
3. Data inconsistency	c) refers to whether the same data kept at different places do or do not match.
4. Data integrity	d) the accuracy and consistency of data across its entire life cycle, from when it is captured and stored to when it is processed, analyzed and used.

5. Data completeness	e) refers to the extent to which all required and expected data elements are present in a dataset. It assesses whether a dataset contains all the necessary information that it is supposed to have.
6. Data analysis	f) the process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making.
7. Customization	g) refers to the action of altering a product or service to suit a person's or company's preferences or requirements; to modify something for a specific task, focusing either on customers or functions.
8. Open source tools	h) code libraries, applications, and other software that are publicly available for anyone to use and modify; allow everyone access to source code, meaning they can make changes or additions to the project as they see fit
9. Metadata	i) the data that describes other data. It can be used to identify, locate and describe digital objects, such as files, images, videos, and websites.

10. An outlier	j) a value or point that differs substantially from the rest of the data.
11. Data Manipulation	k) refers to the process of adjusting, changing, or controlling data to achieve a specific outcome. It essentially involves data modeling, transformation, cleaning, and enrichment to meet business requirements.
12. Scalability	l) the ability of a system, network, or process to grow in size and capabilities by handling more workloads or accommodating more users without compromising performance.
13. Risk mitigation	m) the process of reducing the probability and/or severity of potential losses.

IV. Formulate the following word combinations and statements from the text in another way.

1. … to enhance customer service, facilitate mergers and acquisitions
2. Organizations can avoid costly errors during data analysis.
3. High-quality data meets the specific needs of an organization and adheres to defined standards.
4. Data consistency ensures that data remains uniform across different datasets and systems.
5. Each source offers unique insights and contributes to a comprehensive dataset.
6. Web scraping extracts data from websites.

7. Creating data profiles involves summarizing key characteristics of the dataset.
8. This discovery prompts necessary corrections to maintain data integrity.
9. Pattern recognition involves identifying recurring sequences or trends within the data.
10. Open-source tools allow organizations to tailor their data profiling processes to specific needs.
11. Commercial tools streamline the data profiling process by automating many tasks.
12. data-driven insights
13. Hiring skilled professionals adds to the cost.
14. Adaptability helps address emerging data quality issues.

V. Answer the following questions.

1. What is Data Profiling?
2. Why is Data Profiling important?
3. What are the key concepts of data profiling?
4. What does the process of data profiling begin with?
5. What methods of Data Collection are presented in the text?
6. What techniques are used for analyzing data?
7. What advantages do open source tools provide?
8. What does the process of creating data profiles involve?
9. What is the main idea of data profiles interpretation?
10. What does structure profiling focus on?
11. What are the types of data profiling?
12. What are the functions of content profiling?
13. What techniques and tools for data profiling can be used?
14. What data profiling techniques come with built-in functionalities?
15. What is the contribution of data-driven insights in the enhanced decision-making?
16. How does data profiling help organizations avoid costly errors?
17. Why is it a great challenge to deal with the unstructured data?

18. Why does the process of data profiling involve skilled personnel?
19. Why is it so important to define clear objectives and benchmarks before starting data profiling?
20. What are the benefits of regular monitoring and updating for data profiling?

VI. Speak on the following topics:

1. Key Concepts of data profiling
2. Processes Involved in Data Profiling
3. Types of Data Profiling
4. Techniques and Tools for Data Profiling
5. Benefits of Data Profiling
6. Challenges in Data Profiling
7. Best Practices for Effective Data Profiling

Unit V
MACHINE LEARNING

Vocabulary
circa – (лат.) приблизительно, около
intervention – вмешательство
to credit (to) – приписывать кому-то (авторство)
to coin – создавать
winning chance –шанс на победу
groundwork – основа
reinforcement learning – обучение с подкреплением
repetitively – неоднократно
"goof" button – клавиша «ошибка»
to parse – анализировать
finite automation (automata) – конечный автомат
plethora – изобилие
subset – подгруппа

feasible – осуществимый
convert – преобразовывать
subsequent – последующий
modern-day – современный
leverage –повышать эффективность
iterative – повторяющийся
pivotal – основной

I. Any learning – be it an animal, human or a machine for that matter, begins with an initial set of observations or as we call it – raw data. This kind of data can originate from interactions, transactions, information exchange, examples, experiences, or instructions. A brain – whether it belongs to a human or animal, tries to look for hidden patterns inside that initial data and then uses that processed information to perform further actions like taking decisions, getting values, getting details, distinguish between things and feelings like ***life - threatening events*** vs. safe event, etc.

II. Over time, people tried to devise out methods to implement the same using machine – methods whose primary aim is to allow the computers to learn automatically and enable them to take decisions on our behalf. Right from the early days of Bayes' Theorem in 1763 and its further research done by Pierre-Simon Laplace **circa** 1805, to the Turing's Learning Machine which was proposed by Sir Alan Turing in 1950, huge research has been done to create machines that can learn and become "intelligent". The very first Machine Learning algorithm is **credited to** Arthur Samuel of IBM for his work on programs that can play checkers. These systems and algorithms enabled computers to learn from initial data and that too with no human **intervention** whatsoever.

III. In 1959 Arthur Samuel, an IBM employee and pioneer in the field of computer gaming and artificial intelligence, coined the term "machine learning".

Although the earliest machine learning model was introduced in the 1950s when Arthur Samuel invented a program that calculated the winning chance in checkers for each side, the history of machine learning roots back to decades of human desire and effort to study human cognitive processes. In

1949, Canadian psychologist Donald Hebb published the book “The Organization of Behavior”, in which he introduced a theoretical neural structure formed by certain interactions among nerve cells. Hebb's model of neurons interacting with one another set a groundwork for how AIs and machine learning algorithms work under nodes, or artificial neurons used by computers to communicate data. Other researchers who have studied human ***cognitive systems***[1] contributed to the modern machine learning technologies as well, including logician Walter Pitts and Warren McCulloch, who proposed the early mathematical models of neural networks to come up with algorithms that mirror human thought processes.

IV. By the early 1960s, an experimental "learning machine" with ***punched tape***[2] memory, called Cybertron, had been developed by Raytheon Company to analyze sonar signals, electrocardiograms, and speech patterns using rudimentary reinforcement learning. It was **repetitively** "trained" by a human operator/teacher to recognize patterns and equipped with a **"goof" button** to cause it to reevaluate incorrect decisions. A representative book on research into machine learning during the 1960s was Nilsson's book on Learning Machines, dealing mostly with machine learning for pattern classification. Interest related to pattern recognition continued into the 1970s, as described by Duda and Hart in 1973. In 1981, a report was given on using teaching strategies so that an artificial neural network learns to recognize 40 characters (26 letters, 10 digits, and 4 special symbols) from a computer terminal.

V. In all raw and basic terms, Machine Learning is defined as a set of ***methodologies***[3] that enables systems to automatically learn and improve from various analyses and outputs without being explicitly programmed. To understand what Machine Learning is, we can look at it as the science of making computers to learn and act like a brain does or as humans do, and autonomously improve their learning over time by feeding them data. A machine learning process involves using algorithms **to parse** data, learn from it, and then make a determination or prediction about something in the world without any ***explicit rule-based programming***[4].

VI. The basic building blocks of Machine Learning algorithms involve three important components: Representation, Evaluation, and Optimization.

Representation is the first step of an ML algorithm's implementation where we define a set of classifiers or we define *finite automation* that a computer can understand. Evaluation involves various scoring functions that can represent predictions of either future values or a future outcome and finally, Optimization, which involves a ***Loss/Cost function***[5] that helps in minimizing faults and maximizing efficiency.

The end goal of machine learning algorithms is to make use of the past data, implement each of the above three components and then successfully interpret any new or unseen data – thus proving its worth and might in solving a **plethora** of business problems.

VII. Machine learning, a **subset** of artificial intelligence, is rapidly transforming industries and changing how we live and work. Its ability to learn from data and make predictions or decisions without explicit programming has opened up a world of possibilities. A report by McKinsey estimates that machine learning could add up to $13 trillion to the global economy by 2030. Machine learning applications are becoming more common in our daily lives, from personalized recommendations on *streaming platforms* that keep us engaged to self-driving cars that navigate our roads with greater precision.

VIII. Modern-day machine learning has two objectives. One is to classify data based on models, which have been developed; the other purpose is to make predictions for future outcomes based on these models.

Here are some of ML applications:

1. Image recognition is one of the most common applications of machine learning which makes use of image segmentation techniques. We have done a detailed discussion on the same, where we have seen its wide applications and the concepts around Image Segmentation, and how Machine Learning makes it **feasible**. It is used to identify objects, persons, places, digital images, etc.
2. Speech recognition is a process that involves **converting** voice instructions to text and then perform **subsequent** classification, segmentation, etc. Various virtual assistants like Google Assistant, Siri, Cortana, etc. make use of this technique.
3. Using map-related data such as traffic density, road signs, traffic signs, etc, various applications have been developed that efficiently convey map-related information to users. E.g. Google Maps, Waze, Open Maps, etc.

4. By implementing various Image Segmentation algorithms of Machine Learning, we can perform effective medical diagnostics by recognizing patterns that usually escape the human eye. This has helped in the early detection of tumors, cancers, artery blockages, etc. by implementing ML algorithms on medically generated data.
E.g. A hypothetical algorithm specific to classifying data may use computer vision of moles coupled with ***supervised learning***[5] in order to train it to classify the cancerous moles.
5. Machine translation is a task that generally uses machine learning models developed using highly sophisticated linguistic knowledge and other related data to achieve a correct translation of text from one language to another. Combined with Natural Language Understanding – which also uses Supervised Learning – Machine Translations have become a crucial part of business transactions.
6. Machine learning plays a **pivotal** role in self-driving cars. Tesla, the most popular car manufacturing company is working on self-driving car. It is using unsupervised learning methods, reinforcement learning methods to train car models to detect people and objects.
7. Fraudulent transactions have features and characteristics that separate them from legitimate transactions. Using regression and classification techniques, we can implement the various levels of Fraud Detection, Fraud Reporting, and Fraud Prevention applications.
8. Machine learning's long **short-term memory**[6] neural network is used for the prediction of stock market trends by taking into consideration the fluctuations, patterns, dependent factors, external factors, moving averages, etc.

(https://www.analytixlabs.co.in/blog/what-is-machine-learning/)

Notes:

1. Cognitive systems are computer systems that are designed to understand, interpret, and respond to complex human language and behavior in a way that simulates human cognition. They **leverage** techniques from Artificial Intelligence (AI), machine learning (ML), and natural language processing (NLP) to enable computers to comprehend and interact with humans

more intuitively and effectively. Cognitive systems aim to mimic human cognitive functions such as perception, reasoning, learning, and decision-making, allowing them to perform tasks that were previously challenging for traditional computer systems.

2. **Paper tape** is a sequential data Storage medium consisting of a long, narrow strip of paper with data encoded as punched holes. Each hole represents a Binary bit, with the absence or presence of a hole indicating a 0 or 1, respectively. Paper tape was widely used for data input and output during the early days of computing, from the 1940s to the 1970s. Paper tape is typically 1 inch wide and perforated with a row of holes along each edge. The holes are spaced at regular intervals, and each hole represents a different bit. The holes are punched using a special punch Machine, and the tape can be read using a tape reader. Paper tape has several advantages over other data storage media. It is relatively inexpensive, easy to produce, and can be easily stored and transported. It is also durable and can withstand repeated use. However, paper tape is also relatively slow and can be easily damaged.

3. **The term methodology** is defined as the group of rational mechanisms or procedures, used to achieve an objective, or series of objectives that directs a scientific investigation. This term is directly linked to science, however, the methodology can be presented in other areas such as education, law, etc.

4. **Explicit Programming** provides a useful engineering point balancing modularization and separation in (at least) two cases. First, when a design concept is tightly coupled with particular constructs in a program, separation is unlikely to lead to any benefits of reusability or comprehensibility. Second, concepts that emerge as a system evolves can be encapsulated and recorded, paving the way for later separation when conditions warrant the separation.

5. **Supervised learning** is the act of training the data set to learn by making **iterative** predictions based on the data while adjusting itself to produce the correct outputs. By providing labeled data sets, the model already knows the answer it is trying to predict but doesn’t adjust the process until it produces an independent output.

6. Short-term memory, also known as working memory, refers to the temporary storage of information that is being actively used and manipulated in the mind. It is a cognitive system that allows individuals to hold and process information for a short period of time.

EXERCISES

I. Which of the paragraphs contain the following information?

1. The application of machine learning which makes use of image segmentation techniques
2. The main parts of Machine Learning algorithms
3. The explanation of the end goal of Machine Leaning algorithms
4. The process that involves converting voice instructions to text
5. The complete definition of Machine Learning
6. Machine learning models developed by using highly sophisticated linguistic knowledge
7. The appearance of the first Machine Learning algorithm
8. The significance of Donald Hebb's book "The Organization of Behavior"
9. Machine learning for pattern classification and a special interest related to pattern recognition
10. The content of the first stage of any learning
11. Machine Learning has opened up a great variety of possibilities.

II. Match the synonyms from two columns:

1. circa	a) transform
2. intervention	b) subgroup
3. to credit (to)	c) more than once
4. to coin	d) following
5. groundwork	e) contemporary
6. repetitively	f) repeated

7. to parse	g) realizable
8. plethora	h) improve efficiency
9. subset	i) approximately
10.feasible	j) invent
11. convert	k) abundance
12. subsequent	l) invent
13. modern-day	m) interference
14. leverage	n) foundation
15. iterative	o) analyze

III. Match the following terms with the appropriate definitions

1. Threatening event	a) is a mathematical Model that computes over a finite set of states. … used to recognize patterns and languages. A classic example of it is a state machine that reads a string of symbols and outputs a Yes or no answer to the question of whether the string is accepted by the finite automaton.
2. Artificial intelligence (AI)	b) is a machine learning training method that trains software to make certain desired actions. … based on rewarding desired behaviors and punishing undesired ones.

3. An Artificial Neural Network (ANN)	c) is the situation that has the potential for causing undesirable consequences or impact.
4. Nerve cells	d) is a service that allows users to broadcast live video content over the internet.
5. Reinforcement learning (RL)	e) is technology that enables computers and machines to simulate human learning, comprehension, problem solving, decision making, creativity and autonomy.
6. A finite automaton (FA)	f) is a type of machine learning model inspired by the structure and function of the human brain. It is a network of interconnected nodes, called artificial neurons, that are designed to process and analyze information.
7. Loss function (or Cost function)	g) are the basic functional units of the nervous system, and the adult human brain is thought to contain around 86 billion of them.
8. A streaming platform	h) a foundational element that underpins the efficacy of numerous algorithms and

models. Understanding its significance is crucial to comprehending the intricate workings of AI.

IV. Formulate the following word combinations and statements from the text in another way.

1. These systems and algorithms enabled computers to learn from initial data and that too with no human intervention whatsoever.
2. The history of machine learning roots back to decades of human desire and effort to study human cognitive processes
3. to make a determination or prediction about something in the world without any explicit rule-based programming
4. to make computers to learn and act like a brain does or as humans do, and autonomously improve their learning over time by feeding them data.
5. to interpret any new or unseen data – thus proving its worth and might in solving a plethora of business problems.
6. Streaming platforms keep us engaged to self-driving cars that navigate our roads with greater precision
7. This has helped in the early detection of many diseases by implementing ML algorithms on medically generated data.

V. Answer the following questions.

1. What does any learning begin with?
2. Since when have the scientists been working at developing intelligent machines?
3. What program did Arthur Samuel invent in the 1950s.
4. What was the contribution of Walter Pitts and Warren McCulloch to the modern machine learning technologies?
5. What kind of memory did the experimental leaning machine, called Cybertron, have?
6. What was the principle of Cybertron operation?

7. What can be done in order to understand what Machine Learning is?
8. What ML component includes scoring functions?
9. What is the main idea of the Loss/Cost function?
10. What future does McKinsey predict for Machine Learning?
11. How are traffic predictions implemented?
12. What techniques are used for distinguishing fraudulent transactions from the legitimate ones?
13. What network is used for predicting stock market trends?

VI. Speak on the following topics:

1. The history of machine learning
2. The complete definition of Machine Learning
3. Three important components of the basic building blocks of Machine Learning algorithms
4. The applications of Machine Learning algorithms

Unit VI
MACHINE LEARNING ENGINEER

Vocabulary

simultaneously – одновременно
a ride-hailing application – заказ машины/такси через мобильное приложение
pickup time – время предоставления автомобиля
feature - характеристика
variable – переменная величина
entries – входные данные
counterpart – коллега
GPU – графический процессор
deploy – развертывать
dependency – ресурс, фрагмент кода
standalone – автономный
wraps up – упаковать

monitor – контролировать
destination – конечный пункт
outdated – устаревший
drift – сдвиг, смещение
prerequisite – обязательное условие
responsible – ответственный
background – зд. образование

I. The main idea of a data scientist's job is to make sense of information and answer the question if there are any patterns, which can help us forecast the future. But what if we need to predict the future every day, every hour, every minute and do that for thousands of people **simultaneously**, say, predict the taxi arrival time? There is one specific role that builds the bridge between data science and its practical **counterpart** - machine building. It is an ML engineer, having a solid foundation in computer science, mathematics and statistics, typically acts as a bridge between data scientists who focus on statistical and model-building work and the construction of machine learning and AI systems. So the role of machine learning engineers is to use machine learning to, somehow, bring additional value to the business or the product. The key word here is "product".

II. Let us imagine we have a product team that builds **a ride-hailing application**. What we want is an algorithm that will accurately predict **pickup time** for the customer. We can calculate pickup time based on distance and average time without machine learning, using a simple rule-based system but there are plenty of variables, which may skew the results. Rainfalls and blizzards, traffic congestions and road accidents, all affect the arrival time. With a rule-based system, a software engineer would have to consider all possible factors and write code for them. There are so many of those and there is no way to write rules for everything. How can ml engineers help? They can build a model that learns all the possible relations between data by itself and then give us a more accurate prediction if we support it with the necessary data.

III. ML engineers have some definite responsibilities. An ML engineer starts his work with choosing and preparing data. Let us assume that there

are several variables to calculate pickup time: the distance from the consumer to the driver, speed, weather and traffic congestion to name a few. All of these can become **features**, which a model uses to give us prediction results. To get the data an ml engineer will have to analyze historical records on previous pickups that contain those **variables**. Choosing the right data and consolidating it is the first step in preparation. Then the ml engineer would clean the errors from the data, fill in the missing **entries** and transform records into a single format. Once the data is ready, an ml engineer is to choose the algorithm that would fit the task. The choice depends on the type of data, expected predictive accuracy and how resource - intensive the model is. You may need deep neural networks to process images and videos with 98% accuracy. But training them would require renting clusters of **GPUs[1].** Running these models in production may require specific AI optimized processing units. But, sometimes good old decision trees would be enough. The ml engineer would experiment with several models and a subset of data to find the one that fits the task to start with model training. During the training process, a model will learn predictions by finding patterns in the training data set. You also need a testing set of historical data to evaluate whether the model gives accurate forecasts. If it passes the test, we have a model that can make predictions. But, the model is not a part of the product and the customers cannot use it yet. So, an ml engineer comes to productionalizing the model and its **deployment[2].**

IV. Here is our taxi application or, in this case, two client applications used by drivers and customers and our server where all the back and logic sit. Now we need to deploy the model. Machine learning models are usually deployed as a micro-service, an isolated container where the code has all the **dependencies** and can perform as a **standalone** unit. So, an ml engineer **wraps up** the model into a container and deploys it on the server. Then, he or she needs to connect the model to data sources. The applications will handle some part of the data like driver and customer geolocation, current speed of the car and so on. We'll also need extra data like traffic incidents, jams or weather that comes from a separate database. From this point the model can consume the required data, calculate a prediction and send it back to the customer.

V. There is another problem. We remember that the model was tested on historical data, but how well does it work in real time conditions? We need to track its performance, and this is one of the main concerns of a machine learning engineer: model performance **monitoring** and evaluation. Let's say the model predicted a taxi would arrive in 14 minutes while it actually took 20 minutes. To capture this an ml engineer would set up monitoring infrastructure to compare real world data to the model's predictions in order to understand its accuracy and how it changes over time. Monitoring systems provide ml engineers with necessary data to make a decision whether the model performs well and if it needs retraining. As world conditions are changing, the model can require new data. Say, a large part of a major city highway was closed for reconstruction, which made drivers reach their **destinations** later. The model started predicting a pickup time less accurately because it was trained on **outdated** data and if the ml engineer has monitoring systems set right they will show the **drift**. Such changes are **a prerequisite** for training a new model with fresh data. Since the world conditions may change daily retraining often becomes a daily task for a machine learning engineer.

VI. So, what would the typical **background** and skill set of an ml engineer look like? First, it is statistics, data analysis and applied mathematics. As ml engineers create features and prepare data, the fundamentals are critical. It is obvious that these specialists must also know existing machine learning algorithms and common architectures. Decision trees support **vector machine**s, Naïve Bayes classifiers, deep leaning networks are a few popular algorithms used in ml applications. To train those models engineers have to be familiar with common tools. Python is the main programming language used in data science. Ml engineers may also be proficient in R to explore and visualize data.

A machine-learning engineer is a person in IT who focuses on researching, building and designing self-running artificial intelligence systems to automate predictive models. ML engineers design and create AI algorithms capable of learning and making predictions that define learning. ML engineers typically work within a data science team, collaborating with data scientists,

data analysts, IT experts, DevOps experts, software developers, and data engineers.

(https://www.analytixlabs.co.in/blog/what-is-machine-learning/)

Notes:

1. The GPU, or Graphics Processing Unit, is a specialized electronic circuit designed to rapidly manipulate and alter memory to accelerate the creation of images in a frame buffer intended for output to a display device. In the context of AI, the GPU plays a pivotal role in accelerating the training of machine learning algorithms by performing matrix operations in parallel. With hardware optimized for parallel processing, the GPU excels at performing repetitive tasks by simultaneously executing multiple threads, leading to significant performance improvements in AI-related computations.

2. **Software deployment** includes all of the steps, processes, and activities that are required to make a software system or update available to its intended users. Today, most IT organizations and software developers deploy software updates, patches and new applications with a combination of manual and automated processes. Some of the most common activities of software deployment include software release, installation, testing, deployment and performance monitoring.

EXERCISES

I. Which of the paragraphs contain the following information?

1. One of the main concerns of a machine learning engineer.
2. One of the functions of an ml engineer is to clean the errors from the data.
3. ML engineers must be proficient in some programming languages.
4. The person performing the linkage between the construction of machine learning and AI systems.
5. Sometimes, there occurs the necessity for an ml engineers to apply deep neural networks to process images and videos.

6. There is an explanation of the term “micro-service”
7. The difference between the jobs of a data scientist and a ml engineer
8. An example of Machine learning application.
9. The function of monitory systems.
10. At the stage of training a model, it is necessary to examine historical data.

II. Match the synonyms from two columns:

1. simultaneously	a) shift
2. feature	b) autonomous
3. variable	c) pack
4. entries	d) at a time
5. counterpart	e) obligatory condition
6. standalone	f) obsolete
7. wraps up	g) characteristic
8. monitor	h) input data
9. destination	i) alternating quantity
10. outdated	j) education
11. drift	k) accountable
12. prerequisite	l) colleague
13. responsible	m) terminal
14. background	n) control

IV. Match the following terms with the appropriate definitions

1. GPU cluster	a) a small, loosely coupled service that is designed to perform a specific business function, and each one can be developed, deployed, and scaled independently.

2. processing unit (PU)	b) a series of symbols that separately do not represent anything, but when combined they can generate a language understandable only to those who understand it.
3. decision tree	c) a supervised machine-learning algorithm that is used for classification tasks such as text classification.
4. Micro-service	d) one of the most comprehensive statistical programming languages available, capable of handling everything from data manipulation and visualization to statistical analysis.
5. rule-based system	e) an electronic circuit or chip that performs basic arithmetic and logical operations. It is used in computers, smartphones, and other electronic devices to process data and instructions. It is typically designed to perform specific tasks within a larger system, such as controlling motors, managing software applications, or analyzing data.

6. Code	f) a high-level, general-purpose, interpreted object-oriented programming language, popular among experienced C++ and Java programmers.
7. Naïve Bayes classifier	g) a set of computers where each node is equipped with a Graphics Processing Unit (GPU). Computational demands are ever-rising, whether in cloud or traditional markets
8. Python	h) a supervised machine-learning algorithm that classifies data by finding an optimal line or hyperplane that maximizes the distance between each class in an N-dimensional space.
9. R	i) a non-parametric supervised learning algorithm, which is utilized for both classification and regression tasks.
10. Vector machine (SVM)	j) a type of software system that uses rules as the basis for making decisions or solving problems.

V. Formulate the following word combinations and statements from the text in another way.

1. There are plenty of variables, which may skew the results
2. There are so many of those (factors) and there is no way to write rules for everything.
3. The choice depends on the type of data, expected predictive accuracy and how resource - intensive the model is.
4. Here is … our server where all the back and logic sit.
5. The code has all the dependencies and can perform as a standalone unit
6. To capture this an ml engineer would set up monitoring infrastructure
7. If the ml engineer has monitoring systems set right, they will show the drift.
8. To train those models engineers have to be familiar with common tools.

VI. Answer the following questions.

1. What is the difference between the jobs of a data scientist and an ml engineer?
2. What is the advantage of an algorithm for the prediction as compared to the rule-based system?
3. How does an ml engineer get the data for his job?
4. What actions should an ml engineer take before choosing the right algorithm that would fit the task?
5. How does a machine-learning engineer solve the problem of model performance monitoring and evaluation?
6. Why does the machine learning model need regular retraining?
7. What is the focus of ml engineers' activities?

VII. Speak on the following topics:

1. The role of machine-learning engineers.
2. The examples of ml engineers' activities.
3. ML engineers' responsibilities.
4. The problem of model performance monitoring and evaluation.
ML engineer's background and skills.

Unit VII
ARTIFICIAL INTELLIGENCE

Vocabulary
endow (with) – наделить чем-то
to reason – рассуждать
to generalize –обобщать
proof – доказательство
proficiency – квалификация
domain – область, сфера
attain – приобретать
search engine – поисковая система
trait – отличительная черта
implement – осуществлять
by rote – наизусть
inference – вывод, заключение
premise – зд. данное
tentative model – предварительная модель
irrefutable –неопровержимый
viable – жизнеспособный
biased – необъективный
pledge – обещать
fiscal – финансовый
query –запрос
groundbreaking– инновационный
snippet – отрывок

I. Artificial intelligence (AI) is the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings. The term is frequently applied to the project of developing systems **endowed** with the intellectual processes characteristic of humans, such as the ability **to reason**, discover meaning, **generalize**, or learn from past experience. Since their development in the 1940s, digital computers have been programmed to carry out very complex tasks—such as discovering

proofs for mathematical theorems or playing chess—with great **proficiency**. Despite continuing advances in computer processing speed and memory capacity, there are as yet no programs that can match full human flexibility over wider **domains** or in tasks requiring much everyday knowledge. On the other hand, some programs have **attained** the performance levels of human experts and professionals in executing certain specific tasks, so that artificial intelligence in this limited sense is found in applications as diverse as medical diagnosis, computer **search engines**, voice or handwriting recognition, and **chatbots**[1].

II. Psychologists generally characterize human intelligence not by just one **trait** but by the combination of many diverse abilities. Research in AI has focused chiefly on the following components of intelligence: learning, reasoning, problem solving, perception, and using language.

There are a number of different forms of learning as applied to artificial intelligence. The simplest is learning by trial and error. For example, a simple computer program for solving **mate-in-one**[2] chess problems might try moves at random until mate is found. The program might then store the solution with the position so that, the next time the computer encountered the same position, it would recall the solution. This simple memorizing of individual items and procedures—known as **rote learning**—is relatively easy **to implement** on a computer. More challenging is the problem of implementing what is called generalization. Generalization involves applying past experience to analogous new situations.

III. To reason is to draw **inferences** appropriate to the situation. Inferences are classified as either **deductive**[3] or **inductive.** An example of the former is, "Fred must be in either the museum or the café. He is not in the café; therefore, he is in the museum," and of the latter is, "Previous accidents of this sort were caused by instrument failure. This accident is of the same sort; therefore, it was likely caused by instrument failure." The most significant difference between these forms of reasoning is that in the deductive case, the truth of the **premises** guarantees the truth of the conclusion, whereas in the inductive case, the truth of the premises lends support to the conclusion without giving absolute assurance. Inductive reasoning is com-

mon in science, where data are collected and **tentative models** are developed to describe and predict future behavior—until the appearance of anomalous data forces the model to be revised. Deductive reasoning is common in mathematics and logic, where elaborate structures of **irrefutable** theorems are built up from a small set of basic axioms and rules.

There has been considerable success in programming computers to draw inferences. However, true reasoning involves more than just drawing inferences: it involves drawing inferences relevant to the solution of the particular problem. This is one of the hardest problems confronting AI.

IV. AI research attempts to reach one of three goals: artificial general intelligence (AGI), applied AI, or **cognitive simulation**[5]. AGI (also called strong AI) aims to build machines that think. The ultimate ambition of AGI is to produce a machine whose overall intellectual ability is indistinguishable from that of a human being's. To date, progress has been uneven. Despite advances in large-language models, it is debatable whether AGI can emerge from even more powerful models or if a completely different approach is needed. Indeed, some researchers working in AI's other two branches view AGI as not worth pursuing. Applied AI, also known as advanced information processing, aims to produce commercially **viable** "smart" systems—for example, "expert" medical diagnosis systems and stock-trading systems. Applied AI has enjoyed considerable success. In cognitive simulation, computers are used to test theories about how the human mind works—for example, theories about how people recognize faces or recall memories. Cognitive simulation is already a powerful tool in both neuroscience and cognitive psychology.

V. AI poses certain risks in terms of ethical and socioeconomic consequences. As more tasks become automated, especially in such industries as marketing and health care, many workers are poised to lose their jobs. Although AI may create some new jobs, these may require more technical skills than the jobs AI has replaced.

Moreover, AI has certain biases that are difficult to overcome without proper training. For example, U.S. police departments have begun using predictive policing algorithms to indicate where crimes are most likely to occur. However, such systems are based partly on arrest rates, which are already

disproportionately high in Black communities. This may lead to over-policing in such areas, which further affects these algorithms. As humans are inherently **biased**, algorithms are bound to reflect human biases.

VI. Privacy is another aspect of AI that concerns experts. As AI often involves collecting and processing large amounts of data, there is the risk that this data will be accessed by the wrong people or organizations. With generative AI, it is even possible to manipulate images and create fake profiles. AI can also be used to survey populations and track individuals in public spaces. Experts have implored policymakers to develop practices and policies that maximize the benefits of AI while minimizing the potential risks. In January 2024 singer Taylor Swift was the target of sexually explicit nonconsensual deep fakes that were widely circulated on social media. Many individuals had already faced this type of online abuse (made possible by AI), but Swift's status brought the issue to the forefront of public policy.

VII. LLMs are located at data centers that require large amounts of electricity. In 2020 Microsoft **pledged** that it would be carbon neutral by 2030. In 2024 it announced that in the previous **fiscal year** its carbon emissions had increased by almost 30 percent, mostly from the building materials and hardware required in building more data centers. A **Chat GPT query** requires about 10 times more electricity than a Google Search. **Goldman Sachs** has estimated that data centers will use about 8 percent of U.S. electricity in 2030.

VIII. As of 2024 there are few laws regulating AI. Existing laws such as the European Union's General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) do govern AI models but only insofar as they use personal information. The most wide-reaching regulation is the EU's AI Act, which passed in March 2024. Under the AI Act, models that perform social scoring of citizens' behavior and characteristics and that attempt to manipulate users' behavior are banned. AI models that deal with "high-risk" subjects, such as law enforcement and infrastructure, must be registered in an EU database. AI has also led to issues concerning copyright law and policy. In 2023 the U.S. government Copyright Office began an initiative to investigate the issue of AI using copyrighted works to generate content. That year almost 15 new cases of copyright-related suits were filed

against companies involved in creating generative AI programs. One prominent company, Stability AI, came under fire for using unlicensed images to generate new content. **Getty Images**, which filed the suit, added its own AI feature to its platform, partially in response to the host of services that offer "stolen imagery." There are also questions of whether work created by AI is worthy of a copyright label. Currently, AI-made content cannot be copyrighted, but there are arguments for and against copyrighting it.

IX. Although many AI companies claim that their content does not require human labor, in many cases, such **"groundbreaking"** technology is reliant on exploited workers from developing countries. For example, a Time magazine investigation found that Open AI had used Kenyan workers (who had been paid less than $2 an hour) to sort through text **snippets** in order to help remove toxic and sexually explicit language from Chat GPT. The project was cancelled in February 2022 because of how traumatic the task was for workers. Although Amazon had marketed its **Amazon Go** cashier-less stores as being fully automated (e.g., its AI could detect the items in a customer's basket), it was revealed, that the **"Just Walk Out" technology** was actually powered by outsourced labor from India. More than a thousand workers operated as "remote cashiers," leading to the joke that, in this case, AI stood for Actually Indians.

(https://builtin.com/artificial-intelligence)

Notes:

1. **A chatbot** is a computer program that simulates human conversation through voice commands or text chats or both. Chatbot, short for chatterbot, is an artificial intelligence (AI) feature, that can be embedded and used through any major messaging application.
2. **Mate in One** is a term used to describe a position on a Chessboard that allows the player to move to give checkmate in one move without any previous forced checks or forced moves.
3. **Deductive thinking** is based on logic and the facts. A good detective can use deductive reasoning to find a killer.

4. **Inductive** is a way to describe something that leads to something else, so when applied to reasoning it just means you collect information and draw conclusions from what you observe.
5. **Cognitive simulation** is a method within artificial intelligence and computational neuroscience where computer models are designed to mimic human thought processes, including perception, reasoning, and learning. This technique allows researchers to analyze and understand the complexities of the human brain by testing various hypotheses through virtual environments. By leveraging cognitive simulation, scientists aim to advance areas like human-computer interaction, decision-making, and cognitive development.
6. **Chat GPT** is a generative artificial intelligence chat bot[2][3] developed by Open AI and launched in 2022. It is currently based on the GPT-4o large language model (LLM). Chat GPT can generate human-like conversational responses and enables users to refine and steer a conversation towards a desired length, format, style, level of detail, and language.
7. **The Goldman Sachs Group, Inc.** is an American multinational investment bank and financial services company. Founded in 1869, Goldman Sachs is headquartered in Lower Manhattan in New York City, with regional headquarters in many international financial centers.
8. **Getty Images Holdings, Inc.** (stylized as gettyimages) is a visual media company and supplier of stock images, editorial photography, video, and music for business and consumers, with a library of over 477 million assets.
9. **Amazon Go** is a chain of convenience stores in the United States and the United Kingdom, operated by the online retailer Amazon. The stores are cashierless.
10. **Just Walk Out technology** allowed customers to grab grocery items from a shelf and walk out of the store.

EXERCISES

I. Which of the paragraphs contain the following information?

1. Classification of the inferences.

2. The law regulation of Artificial Intelligence
3. A considerable success of Applied AI
4. The introduction of the cashier-less technology.
5. The issue of AI using copyrighted works to generate content
6. The promise to decrease carbon emissions
7. Online abuse is, possibly made by AI.
8. The examples of deductive and inductive inferences.
9. The difference between two forms of reasoning: the deductive and inductive ones.
10. A challenging problem of implementing generalization.
11. Some facts from the history of computers.
12. Three goals of AI research
13. The use of AI by the police.

II. Match the synonyms from two columns:

1. to reason	a) evidence
2. to generalize	b) distinctive feature
3. proof	c) subjective
4. proficiency	d) by heart
5. domain	e) to reflect
6. attain	f) capable of living
7. trait	g) summarize
8. implement	h) promise
9. by rote	i) competence
10. inference	j) financial
11. premise	k) financial
12. irrefutable	l) request
13. viable	m) innovative
14. biased	n) indisputable
15. pledge	o) realize
16. fiscal	p) fragment
17. query	q) sphere
18. groundbreaking	r) data
19. snippet	s) gain

III. Match the following terms with the appropriate definitions

1. Rote learning	a) is a statement assumed to be true to start a new argument or theory. It is considered the starting point of reasoning and proof. The word itself originated from the Greek meaning "to be worthy" and is regarded as a universal truth in terms of mathematics
2. Intelligence	b) is the representation of generalized human cognitive abilities in software so that, faced with an unfamiliar task, its system could find a solution. Its intention is to perform any task that a human being is capable of.
3. Proficiency	c) is the repetition of learning material in order for it to be committed to long-term memory.
4. A search engine	d) is a branch of artificial intelligence focused on the practical implementation of AI technologies in real-world scenarios.

5. An inference	e) is a term used in psychology to describe anything related to thinking, learning, and understanding.
6. Failure	f) is the ability to acquire, understand, and use knowledge
7. An axiom	g) is a branch of artificial intelligence focused on the practical implementation of AI technologies in real-world scenarios.
8. Artificial general intelligence (AGI)	h) refers to a high level of skill or expertise in a particular subject or activity. It is the ability to perform a task with competence and excellence, often surpassing the average level of proficiency.
9. Applied AI	i) is a type of software designed to help you find specific information online. It does this by methodically searching through web content based on the specific keywords a user enters into the search box.

10. **A large language model**	j) is a mental process by which individuals draw conclusions from available information. It is a fundamental aspect of human reasoning, allowing us to make sense of the world around us.
11. **Cognitive**	k) is the social concept of not meeting a desirable or intended objective, and is usually viewed as the opposite of success

IV. Formulate the following word combinations and statements from the text in another way.

1. The project of developing systems endowed with the intellectual processes characteristic of humans
2. To reason is to draw inferences appropriate to the situation.
2. There are as yet no programs that can match full human flexibility over wider domains
3. Generalization involves applying past experience to analogous new situations.
4. Elaborate structures of irrefutable theorems are built up from a small set of basic axioms and rules.
5. To date, progress has been uneven.
6. As humans are inherently biased, algorithms are bound to reflect human biases.
7. … sexually explicit non-consensual deep fakes
8. In 2020, Microsoft pledged that it would be carbon neutral by 2030.
9. … models that perform social scoring of citizens' behavior and characteristics

10. One prominent company, Stability AI, came under fire for using unlicensed images to generate new content.
11. “groundbreaking” technology is reliant on exploited workers
12. The project was canceled in February 2022 because of how traumatic the task was for workers.

V. Answer the following questions

1. What is the definition of artificial intelligence?
2. How do psychologists generally characterize human intelligence?
3. What is the difference between deductive and inductive forms of reasoning?
4. What are tentative models developed for?
5. What is the ultimate ambition of AGI?
6. What is the applied AI aimed at?
7. What kind of risks does artificial intelligence face?
8. What benefits do the police obtain when using AI?
9. What is the reason for the wrong assessment of data?
10. How does Microsoft explain the increase of carbon emissions?
11. What is the most wide-reaching law regulation of AI?
12. What AI models should be registered in an EU database?
13. What is the main idea of a copyright law issued in 2023?
14. What did Open AI use Kenyan worker for?
15. Why was the project concerning “groundbreaking” technology cancelled?

VI. Speak on the following topics:

1. The main goals of artificial intelligence
2. Different forms of learning as applied to artificial intelligence
3. Deductive and inductive inferences
4. Three goals of artificial intelligence
5. Risks facing artificial intelligence
6. Law regulation of Artificial Intelligence

Unit VIII
WHAT IS CYBERSECURITY?

Vocabulary
mitigate – уменьшить ransom – выкуп (redeem)
malware – вредоносное ПО
extortion – вымогательство
outage – сбой
pervasive – распространенный
hijack – похищать
malicious – вредоносный
unauthorized – несанкционированный
flaw – слабое место
wearable – портативный
disrupt – разрушать
hostage – заложник
trick – обмануть
trick – афера, жульничество
credential theft – кража учетных данных
earmark – отличительный признак

I. Cybersecurity refers to any technologies, practices and policies for preventing cyberattacks or **mitigating** their impact. Cybersecurity aims to protect computer systems, applications, devices, data, financial assets and people against **ransom**ware and other malware, phishing scams, data theft and other cyber threats.

II. At the enterprise level, cybersecurity is a key component of an organization's overall risk management strategy. According to **Cybersecurity Ventures**[1], global spending on cybersecurity products and services will exceed USD 1.75 trillion total during the years 2021 through 2025. Cybersecurity is important because cyberattacks and cybercrime have the power to disrupt, damage or destroy businesses, communities and lives. Successful cyberattacks lead to identity theft, personal and corporate **extortion**, loss of

sensitive information and business-critical data, temporary business **outages**, lost business and lost customers and, in some cases, business closures. Cyberattacks have an enormous and growing impact on businesses and the economy. By one estimate, cybercrime will cost the world economy USD 10.5 trillion per year by 2025. The cost of cyberattacks continues to rise as cybercriminals become more sophisticated.

III. Apart from the sheer volume of cyberattacks, one of the biggest challenges for cybersecurity professionals is the ever-evolving nature of the information technology (IT) landscape, and the way threats evolve with it. Many emerging technologies that offer tremendous new advantages for businesses and individuals also present new opportunities for threat actors and cybercriminals to launch increasingly sophisticated attacks. For example:

The **pervasive** adoption of cloud computing can increase network management complexity and raise the risk of cloud misconfigurations, improperly secured APIs and other avenues hackers can exploit. More remote work, hybrid work and **bring-your-own-device (BYOD)**[2] policies mean more connections, devices, applications and data for security teams to protect.

Proliferating Internet of Things (IoT) and connected devices, many of which are unsecured or improperly secured by default, can be easily **hijacked** by bad actors. The rise of artificial intelligence (AI), and of generative AI in particular, presents an entirely new threat landscape that hackers are already exploiting through prompt injection and other techniques.

IV. Comprehensive cybersecurity strategies protect all of an organization's IT infrastructure layers against cyber threats and cybercrime. Some of the most important cybersecurity domains include: AI security, Critical infrastructure security, Network security, Endpoint security, Application security, Cloud security Information security, Mobile security.

Here they are:

1. AI security refers to measures and technology aimed at preventing or mitigating cyber threats and cyberattacks that target AI applications or systems or that use AI in **malicious** ways. Generative AI offers threat actors new attack vectors to exploit. Hackers can use malicious prompts to manipulate AI apps, poison data sources to distort AI outputs and even trick AI tools into sharing sensitive information. They can also use (and have already used)

generative AI to create malicious code and phishing emails. AI security uses specialized risk management frameworks—and increasingly, AI-enabled cybersecurity tools—to protect the AI **attack surface**.

2. Critical infrastructure security protects the computer systems, applications, networks, data and digital assets that a society depends on for national security, economic health and public safety.

3. Network security focuses on preventing **unauthorized** access to networks and network resources. It also helps ensure that authorized users have secure and reliable access to the resources and assets they need to do their jobs.

4. Application security helps prevent unauthorized access to and use of apps and related data. It also helps identify and mitigate **flaws** or vulnerabilities in application design. Modern application development methods such as DevOps and **DevSecOps**[3] build security and security testing into the development process.

5. Cloud securit*y* secures an organization's cloud-based services and assets, including applications, data, virtual servers and other infrastructure. Generally speaking, cloud security operates on the shared responsibility model. The cloud provider is responsible for securing the services that they deliver and the infrastructure that delivers them. The customer is responsible for protecting their data, code and other assets they store or run in the cloud.

6. Information security (InfoSec) protects an organization's important information—digital files and data, paper documents, physical media—against unauthorized access, use or alteration.

7. Data security, the protection of digital information, is a subset of information security and the focus of most cybersecurity-related InfoSec measures.

8. Mobile security encompasses cybersecurity tools and practices specific to smartphones and other mobile devices, including mobile application management (MAM) and enterprise mobility management (EMM).

More recently, organizations are adopting **unified endpoint management (UEM)**[4] solutions that allow them to protect, configure and manage all endpoint devices, including mobile devices, tablets, **wearables** and more**.**

V. Some of the common cybersecurity threats are the following: Malware, Ransomware, Phishing, Credential theft and abuse, Insider threats, AI

attacks, Cryptojacking, Distributed denial of service (DDoS). Hackers and cybercriminals create and use malware to gain unauthorized access to computer systems and sensitive data, hijack computer systems and operate them remotely, **disrupt** or damage computer systems, or hold data or systems hostage for large sums of money.

Ransomware is a type of malware that encrypts a victim's data or device and threatens to keep it encrypted—or worse—unless the victim pays a ransom to the attacker. The earliest ransomware attacks demanded a ransom in exchange for the encryption key required to unlock the victim's data. Starting around 2019, almost all ransomware attacks were double extortion attacks that also threatened to publicly share victims' data; some triple extortion attacks added the threat of a distributed denial-of-service (DDoS) attack. In the meantime, ransomware attackers have repurposed their resources to start other types of cyber threats, including info-stealer malware that allows attackers to steal data and hold it **hostage** without locking down the victim's systems and data destruction attacks that destroy or threaten to destroy data for specific purposes.

VI. Phishing attacks are email, text or voice messages that **trick** users into downloading malware, sharing sensitive information or sending funds to the wrong people. Most users are familiar with bulk **phishing scams**[5] — mass-mailed fraudulent messages that appear to be from a large and trusted brand, asking recipients to reset their passwords or reenter credit card information. More sophisticated phishing scams, such as **spear phishing**[6] and business email compromise (BEC), target specific individuals or groups to steal especially valuable data or large sums of money. Phishing is just one type of social engineering, a class of "human hacking" tactics and interactive attacks that use psychological manipulation to pressure people into taking unwise actions.

VII. The X-Force Threat Intelligence Index[7] found that identity-based attacks, which hijack legitimate user accounts and abuse their privileges, account for 30% of attacks. This makes identity-based attacks the most common entry point into corporate networks. Hackers have many techniques for stealing credentials and taking over accounts.

Insider threats are threats that originate with authorized users—employees, contractors, business partners—who intentionally or accidentally misuse their legitimate access or have their accounts hijacked by cybercriminals. Insider threats can be harder to detect than external threats because they have the **earmarks** of authorized activity and are invisible to antivirus software, firewalls and other security solutions that block external attacks.

VIII. Much like cybersecurity professionals are using AI to strengthen their defenses, cybercriminals are using AI to conduct advanced attacks. In generative AI fraud, scammers use generative AI to produce fake emails, applications and other business documents to fool people into sharing sensitive data or sending money.

The X-Force Threat Intelligence Index reports that scammers can use open source generative AI tools to craft convincing phishing emails in as little as five minutes. For comparison, it takes scammers 16 hours to come up with the same message manually. Hackers are also using organizations' AI tools as attack vectors. For example, in prompt injection attacks, threat actors use malicious inputs to manipulate generative AI systems into leaking sensitive data, spreading misinformation or worse.

IX. Crypto-jacking happens when hackers gain access to an endpoint device and secretly use its computing resources to mine cryptocurrencies such as bitcoin, ether or **monero**[8]. Security analysts identified crypto jacking as a cyber-threat around 2011, shortly after the introduction of cryptocurrency. According to the IBM X-Force Threat Intelligence Index, crypto jacking is now among the top three areas of operations for cybercriminals.

(https://www.techgeekbuzz.com/blog/what-is-cybersecurity/)

Notes:

1. Cybersecurity Ventures is a venture capital firm that invests in cybersecurity startups. The company was founded by Mike McNerney and John Viega.

Venture capital (VC) is a form of private equity and a type of financing for startup companies and small businesses with long-term growth potential.

2. BYOD, or bring your own device, refers to corporate IT policy that determines when and how employees, contractors and other authorized end users can use their own laptops, smartphones and other personal devices on the company network to access corporate data and perform their job duties.
3. DevSecOps, which is short for development, security and operations, is an application development practice that automates the integration of security and security practices at every phase of the software development lifecycle, from initial design through integration, testing, delivery and deployment.
4. UEM, or unified endpoint management, is software that enables IT and security teams to monitor, manage and secure all of an organization's end-user devices, such as desktops and laptops, smartphones, tablets, **wearables** and more, in a consistent manner with a single tool, regardless of operating system or location.
5. a phishing scam is a type of cybercrime in which criminals send emails or messages that appear to come from a legitimate source such as a bank, government agency, or payment processing service. They then use the email or message to persuade the recipient to disclose confidential information. This information can include account numbers, passwords, or credit card information.
6. Spear phishing is a malicious email spoofing attack that targets a specific organization or individual, seeking unauthorized access to sensitive information. Spear phishing attempts are not typically initiated by random hackers, but are more likely to be conducted by perpetrators out for financial gain, trade secrets or military information.
7. The X-Force Threat Intelligence Index the annual report of IBM X-Force analysts, which describes changes in the field of cyber threats.
8. Monero (XMR) is an efficient cryptocurrency developed on Blockchain technology to be open-source, allowing anyone to use it, created to achieve privacy while performing digital assets transactions.

EXERCISES

I. Which of the paragraphs contain the following information?

1. The kind of cybersecurity threats demanding redeem.
2. Solutions that allow to protect, configure and manage all endpoint devices
3. A class of attacks, which has as a target employees having an access to company funds.
4. The most common entry point into corporate networks
5. The top three areas of operations for cybercriminals.
6. Some changes in the activity of ransomware attackers
7. The use of generative AI tools for good and bad.
8. The threats that originate with authorized users
9. Cybersecurity aims
10. Cybersecurity products and services need huge funds
11. The cybersecurity strategy applied for national security, economic health and public safety.
12. Generative AI offers hackers new attack vectors to exploit
13. The cybersecurity strategy applied for national security, economic health and public safety.
14. Negative consequences of cyberattacks and cybercrime.

II. Match the synonyms from two columns:

1. mitigate	a) steal
2. ransom	b) unsanctioned
3. outage	c) portable
4. pervasive	d) reduce
5. hijack	e) cheat
6. malicious	f) fraud
7. unauthorized	g) harmful
8. flaw	h) redeem
9. wearable	i) destroy
10. disrupt	j) distinguishing feature

11. to trick	k) weakness
12. trick	l) malfunction
13. earmark	m) widespread

III. Match the following terms with the appropriate definitions

1. Cloud computing	**a)** refers to all the possible points, also called attack vectors, where cybercriminals can access a system and steal data.
2. The Internet of Things (IoT)	**b)** is a type of malware that gathers sensitive information stored on a device. Once a computer has been infected, the info-stealer uses various techniques to acquire data.
3. Prompt injection	**c)** is a psychological manipulation technique used by threat actors to get others to do things or reveal private information.
4. An attack surface	**d)** is a type of cyber-attack where the attacker conceals the original identity and pretends to be a trusted and authorized one to gain access to a computer or network.

5. Data security	**e)** refers to a network of physical devices, vehicles, appliances, and other physical objects that are embedded with sensors, software, and network connectivity, allowing them to collect and share data.
6. Malware	**f)** is a security exploit in which the attacker targets an employee who has access to company funds and convinces the victim to tranfer money into a bank account controlled by the attacker.
7. An information stealer	**g)** is the on-demand delivery of computing services such as servers, storage, databases, networking, software, and analytics. Rather than keeping files on a proprietary hard drive or local storage device, it makes possible to save remotely.
8. A spoofing attack	**h)** is a type of attack or manipulation technique used primarily in the context of large language models (LLMs) and natural language processing (NLP) systems

9. Business email compromise (BEC)	**i)** is a subset of information security and the focus of most cybersecurity-related InfoSec measures.
10. Social engineering	**j)** is any software code or computer program that is intentionally written to harm a computer system or its users. Almost every modern cyberattack involves some type of malware.

IV. Formulate the following word combinations and statements from the text in another way.

1. By one estimate, cybercrime will cost the world economy USD 10.5 trillion per year by 2025.
2. Apart from the sheer volume of cyberattacks, one of the biggest challenges for cybersecurity professionals is the ever-evolving nature of the information technology (IT) landscape, and the way threats evolve with it.
3. Mobile security encompasses cybersecurity tools and practices specific to smartphones and other mobile devices.
4. … to fool people into sharing sensitive data or sending money.

V. Answer the following questions.

1. What is cybersecurity aimed at?
2. Why do organizations consider cybersecurity to be the key component of their overall risk management strategy?
3. Why do the expenses for cybersecurity continue rising?
4. How can hackers use generative IA and malicious prompts?
5. What are DevOps and DevSecOps used for?
6. What cybersecurity strategy operates on the shared responsibility model?

7. What kind of threats does ransomware presume?
8. What kind of cyberattacks make users download malware?
9. Why are insider threats harder to detect than external threats?
10. When does crypto jacking happen?

VI. Speak on the following topics:

1. Cybersecurity main functions
2. The results of cyberattacks and cybercrime
3. Challenges of cybersecurity
4. Comprehensive cybersecurity strategies
5. The common cybersecurity threats

Unit IX
CYBERSECURITY ENGINEER

Vocabulary
cryptographic algorithms – алгоритм шифрования информации
address vulnerabilities - устранять уязвимости
ransom - выкуп
ransomware – программа-шантажист; программа-вымогатель
penetrating testing – тестирование на проникновение
trusted system – защищенная (безопасная) система
security breach – нарушение безопасности
ensure - гарантировать, обеспечивать
security posture – средства обеспечения безопасности
plentiful – в избытке
smooth functioning - бесперебойное функционирование

I. Cybersecurity job growth is robust. The US Bureau of Labor Statistics projects that “employment of information security analysts is projected to grow 32% from 2022 to 2032, faster than the average for all occupations.”

A cybersecurity engineer is a specialist who applies engineering principles and practices in the process of designing, developing, implementing and managing of secure computer systems. This vocation assumes a wide range of activities. Among them are the design of secure networks and authentication systems. Cybersecurity engineers write Secure Code and **cryptographic algorithms**. They also analyze risks concerning computer system security and **implement** security controls. These specialists deal with system vulnerabilities developing tests for their detection.

II. Thus, cybersecurity engineers must be able to detect and neutralize the threats and vulnerabilities in systems and software. It is very important for them to possess the skills to develop and implement high-tech solutions in order **to address vulnerabilities,** to defend against hacking, malware, ransomware. Cybersecurity engineers are specialists capable to supply consumers with the defense against insider threats and all kinds of cybercrimes.

III. Taking into account that individuals and business are facing advanced persistent threats to computer systems, cybersecurity experts, also called information security engineers, are of great demand for the economy and science in any country. They are usually highly qualified specialists being able to perform assessments, penetration testing and develop trusted systems. Among cybersecurity engineer responsibilities (in addition to those mentioned above) are such as planning, implementing, managing, monitoring and upgrading security measures for the protection of the organization's data, systems and networks; responding to all system and/or network security breaches; ensuring that the organization's data and infrastructure are protected by enabling the appropriate security controls.

IV. Job opportunities in the cybersecurity engineering field are plentiful. Graduates with the appropriate educational background can be offered such positions as network security engineer, IT security engineer, information security engineer, information assurance engineer, information systems security engineer. People with cybersecurity engineering skills and experience may also want to explore or position themselves for such connected positions as: cybersecurity architect, cybersecurity manager, cybersecurity consultant, cybersecurity director, chief information security officer (CISO).

V. Thus, the technological boom results in the growing need for the cybersecurity professionals. They can guarantee smooth functioning of all areas of our life: industry, transport, business, education, culture etc.

(https://www.mygreatlearning.com/blog/information-security-engineer/)

EXERCISES

I. Which of the paragraphs contain the following information?

1. Great demand for cybersecurity engineers
2. The direct connection between the number of cybersecurity engineers and technological revolution
3. The definition of a cybersecurity engineer
4. Cybersecurity engineers' skills
5. Job offers for cybersecurity engineers
6. Additional responsibilities of a cybersecurity engineer
7. Cybersecurity engineers' main functions

II. Match the synonyms in every line

1.	authentication	a) profession	b) identification
2.	design	a) pattern	b) project
3.	analyze	a) renew	b) study
4.	security	a) confidence	b) safety
5.	detect	a) guarantee	b) reveal
6.	malware	a) weak spot	b) malicious software
7.	authentication	a) violation	b) recognition
8.	algorithm	a) project	b) pattern
9.	vocation	a) client	b) profession
10.	trust	a) recognition	b) safety
11.	ensure	a)introduce	b) guarantee
12.	vulnerability	a)malicious software	b) weak spot
13.	breach	a) safety	b) violation
14.	upgrade	a) study	b) renew
15.	customer	a) identification	b) client

III. Match the following terms with the appropriate definitions

1. Ransomware	**a)** designed to serve the purpose of providing security. Safety is ensured by protecting the system against malicious software's and third party intruders, allowing only verified users to access the computer system; responsible for providing security at different levels and based on different parameters
2. Security Code	**b)** any incident that results in unauthorized access of data applications, services, networks and/or devices by bypassing their underlying security mechanisms; occurs when an individual or an application illegitimately enters a private, confidential or unauthorized logical IT perimeter
3. Trusted system	**c)** a type of malicious software, or malware, that threatens a victim by destroying or blocking a access to critical data or systems until a ransom is paid. It used to target individuals

	but now targets organizations and has become a larger and more difficult threat to prevent and reverse
4. Penetration test	**d)** … a series of numbers that, in addition to the bank card number printed on a card. … is used as a security feature for card not present transactions, where a personal identification number (PIN) cannot be manually entered by the cardholder (as they would during point-of-sale or card present transactions).
5. Security breach	**e)** the procedure of scrutinizing an IT foundation's security; a cybersecurity technique that organizations use to identify, test and highlight vulnerabilities in their security posture; often carried out by ethical hackers

IV. Formulate the following statements from the text in another way.

1. This vocation assumes a wide range of activities.
2. They also analyze risks concerning computer system security and implement security controls.

3. Cybersecurity engineers must be able to detect and neutralize the threats and vulnerabilities in systems and software.
4. It is very important to possess the skills … to address vulnerabilities.
5. ... individuals and business are facing advanced persistent threats to computer systems
6. Job opportunities in the cybersecurity engineering field are plentiful.
7. The technological boom results in the growing need for the cybersecurity professionals.

V. Answer the following questions:

1. How is the profession of a cybersecurity engineer defined?
2. What activities does the vocation of a cybersecurity engineer assume?
3. What abilities must a cybersecurity engineer have?
4. Why is there a great need for cybersecurity engineers?
5. What knowledge and skills must cybersecurity engineers possess?
6. What positions can specialists in cybersecurity engineering hold?
7. What do we mean by the connected positions for the experts in cybersecurity engineering?

VI. Speak on the following topics

1. Cybersecurity engineers' skills
2. Responsibilities of a cybersecurity engineer
3. Cybersecurity engineers' main functions

SUPPLEMENTARY READING

Read the following texts and be ready to answer the questions given after

WHAT ARE NEURAL NETWORKS AND WHAT DO WE NEED THEM FOR?

For a long time, people have been thinking on how to create a computer that could think like a person. The advent of artificial neural networks is a significant step in this direction. Our brain consists of neurons that receive information from sensory organs and process it: we recognize people we know by their faces, and we feel hungry when we see delicious food. All of this is the result of brain neurons working and interacting with each other. This is also the principle that artificial neural networks are based on, simulating the processes occurring in the human brain.

Artificial neural networks are a software code that imitates the work of a brain and is capable of self-learning. Like a biological network, an artificial network also consists of neurons, but they have a simpler structure. If you connect neurons into a sufficiently large network with controlled interaction, they will be able to perform quite complex tasks. For example, determining what is shown in a picture, or independently creating a photorealistic image based on a text description.

How a neural network works

An artificial neuron receives signals through several inputs, then transforms them and transmits them to other neural. That is, the work of a neuron is to convert several parameters into one.

For instance, a neural network is trying to determine if there is an image of a tabby cat in a picture. It has already processed hundreds of thousands of cat photos and knows that the color of their coat is a combination of certain shades, such as black, silver and brindle. Signals are sent to the neuron that these three colors are predominant in the image, which means that, most likely, there is a cat in it. Next, the neural network checks if the picture has eyes, ears, and a tail. If all four factors match, it can state with confidence that the image is that of a cat.

In a very simplified way, the operational scheme of a neural network is like this. Imagine that for each “yes” answer we get 1 point, and for “no” we get 0 points. If, after verification, the neural network gets 4 points, it is totally confident that there is an image of a cat in the picture. If the result is 2-3 points, there is a high probability that the cat is there, but it might have hidden its tail. If the result is 1-0 points, then there is definitely no cat in the picture. Or it has hidden well.

What do we need neural networks for?

Neural networks help people to get rid of monotonous routine work by doing it much faster for them. For example, unmanned vehicles are being developed based on neural networks, which can free drivers from actual driving in the near future. People will be able to work, study or have fun on the road instead of monitoring the traffic situation. Neural networks in the Moscow metro are used when paying the ticket fee through one’s biometric data and when searching for people who are on the federal wanted list.

Neural networks are also used for forecasting, image recognition, controlling, recognition of hidden patterns in a large amount of data, as well as for solving tasks related to artificial intelligence, machine and deep learning.

The tasks that neural networks solve

Depending on the task that a neural network performs, it can be classified into one of the five types.

Classification

A neural network recognizes a person by their face or determines what is shown in the picture. For example, the Google neural net can identify what you are drawing with your mouse on a canvas.

Regression

A neural network can predict growth of stocks, the value of real estate or a person’s age based on a photo. A neural net determining gender based on a photo will help to avoid mistakes when filling out documents

Time series forecasting

A neural network makes forecasts of weather, price rises or electricity consumption. This also includes neural networks that control unmanned vehicles. They predict the behavior of other road users based on analyzing millions of hours of dashcam recordings.

Clustering

A neural network combines large amounts of data into groups according to certain criteria. The DeepCluster neural network arranges photos by subject: sunsets, planes, forest and buses.

Generation

A neural network creates music, images, video or text according to the given parameters. The Yandex neural net not only creates poems from user search queries, but also reads them aloud.

How do they train neural networks?

The main advantage of neural networks is their ability to self-learn. If we return to the example of finding a cat in a photo, then, after confusing it several times with a fox, a neural network will conclude that pointed ears is not the most specific attribute of a cat. And then it will start giving not 1, but 0.5 points for the answer "yes". A well-trained neural network can recognize data that was not in the training set, as well as corrupt or incomplete data. For example, it will recognize a cat in a photo, even if only part of its face is visible.

(https://habr.com/en/articles/676266/)

Questions:

1. What principle are artificial neural networks based on?
2. What is the difference between artificial neural networks and a biological network?
3. How does a neural network work?
4. What are the reasons of our need for neuron networks?
5. What tasks do neuron networks perform?
6. How are neuron networks trained?

A TEXT NEURO EDITOR AND OTHER PRODUCTIVITY TOOLS IN YANDEX BROWSER

Yandex Browser has added productivity tools based on YandexGPT technology. Now the Browser helps to effectively solve everyday tasks related to content, without using separate services or applications.

Neural networks in the editor will help users create texts from scratch and improve ready—made ones - for example, correct errors, rewrite in a certain style or format. The Browser has a built-in translator with YandexGPT, which uses the appropriate vocabulary depending on the subject area of the text, and the tool for short retelling in the computer version can now work with documents in PDF, DOCX, TXT format.

All the tools can be used in a separate interface: you can access it by clicking on the buttons under the search bar on the main browser page. They are available not only in the version for computers and laptops, but also in the mobile version.

The neural editor allows you to create texts using YandexGPT, edit them, and change the style. It can be opened by clicking the "Edit" button next to any site in the browser. The tool will help you prepare a presentation plan or write a speech text, edit an online document, turn a dry informational message into a light text for a social network.

You can change the text using built-in commands (for example, "More formally" or "In Simple words") or at any user's request via chat. For example, you can ask the editor to shorten the text and adapt it for presentation slides, rewrite it in an official business style, or split it into list items. The editor supports Russian and English languages.

You can start working with text on your smartphone and then continue on your computer. The editor has a history of working with texts — for quick access to them. Thanks to this, it is now convenient to write and store your own notes directly in the browser.

To work with content in other languages in Yandex Browser now has access to the beta version of the updated Translator based on YandexGPT.

It learned to distinguish between the subject area of the text and use the appropriate vocabulary. So, the tool will select special terms for a scientific article, and common terms for a post from a culinary blog.

The translator, which understands more than 100 languages, can now upload arbitrary texts, images or links to websites. To do this, simply drag and drop the file or copy the link to the chat. The beta version of the expert Translator is available by clicking on the button under the search bar on the Browser's home page.

In the Browser, you can now briefly retell the downloaded text of any length: it will cope with both a scientific treatise and a voluminous fiction novel. On computers, the function is available not only for texts and videos, but also for documents in popular PDF, DOCX, and TXT formats.

Now there are two retelling modes: short and detailed — with a lot of details from the original. The tool also knows how to break the resulting retelling into semantic chapters and select a subtitle for each one so that it is easier for the user to navigate the text. Yandex integrated a new generation of neural networks into its Browser last, which gave users a number of opportunities — for example, to improve their texts directly in the browser, generate images, translate videos from eight foreign languages into Russian, retell the content of videos and create subtitles for them.
(https://www.akm.ru/eng/press/a-text-neuro-editor-and-other-productivity-tools-have-appeared-in-yandex-browser/)

Questions:

1. What technology are productivity tools of Yandex Browser based on?
2. What are the abilities of Yandex Browser built-in translator?
3. What operations can the neural editor perform with the text?
4. What makes it convenient to write and store our own notes directly in the browser.
5. How many languages does the Browser translator understand?
6. What advantages does the Browser provide for retelling different kinds of information?

ETHICAL CHALLENGES IN WORKING WITH NEURAL NETWORKS

Today, virtually no online business operates without content generated by neural networks. This technology significantly eases the work of specialists: you can enjoy the creative process rather than deal with routine tasks. However, every time we use AI, the following questions come to mind.

1. Who owns the content or code created by a neural network?
2. Do you want to assist your competitors?
3. How to deal with the risk of legal action for using content created by AI?
4. What if there's no longer a need for manually created content in the future?
5. Isn't there a risk in giving confidential and corporate information to neural networks?
6. Who is responsible for incorrect information?
7. Can AI replace humans in the workplace?
8. How will the implementation of AI affect the cost of content?

Generative neural networks are not always accurate. But the main issue arises when sensitive information about a company is uploaded to AI; the AI developer gains access to that information and can freely utilize it. Therefore, confidential data should not be handed over to neural networks without the owner's permission. When training AI models, it's crucial to ensure that the information used for training was obtained legally and doesn't violate any laws or regulations. Most publicly available generative AI services (such as ChatGPT) do not disclose the datasets they were trained on. As a result, it's unclear who actually owns the data.

The situation is exacerbated by the inability to verify the objectivity of the results, and AI bias is a real problem. Robots 'pick up' stereotypes from training data. They can convincingly lie. Therefore, it's unwise to trust content generated by AI as the sole source of information, especially when it comes to scientific research.

Identifying issues with AI is not difficult. Content created solely by neural networks stands out significantly:

- Long, complex paragraphs with repetitive content

- Clumsy clichés, awkward structure, and syntactically long sentences
- Abundant use of adjectives
- Facts based on subjective opinion.

Ethical rules for using artificial intelligence

Using generative AI is recommended with a license from the company. This is important in terms of responsibility and ownership rights. However, it's crucial not to directly copy the content or code of a neural network. Always edit, modify, and supplement material to make it unique.

- Data protection is the most important aspect. The accuracy of information generated by neural networks is questionable, and it should not be blindly trusted, especially in the case of rapidly changing topics.
- The output of a neural network should not be copied verbatim. Generating content is good, but always remember the need to refine the output for truly fresh ideas.
- Artificial intelligence is excellent for surface-level exploration and idea testing. It should not be used to prove one's ideas.
- Most information is considered 'sensitive' content. Any information entered into a neural network may be accessible to competitors, current and potential clients.
- Do not share personal data with AI tools like ChatGPT. The same applies to customers' personal data (names, email addresses, phone numbers, and other personally identifiable information).

(https://tooploox.com/datasets-and-ai-ethics)

Questions

1. What are the concerns about the usage of neural networks?
2. What recommendation is given to the owners of confidential data and why?
3. What are the main ethical rules for those using artificial intelligence?

ЗАКЛЮЧЕНИЕ

Высококвалифицированные специалисты в сфере информационных технологий – это люди, стремящиеся к профессиональному развитию и самообразованию. В своей работе они часто сталкиваются с необходимостью решать нестандартные задачи и в поисках нужной информации обращаются не только к русскоязычным, но и к англоязычным источникам, включая интернет-ресурсы.

Автор выражает надежду, что тексты и упражнения, представленные в издании, позволят учащимся усовершенствовать свои знания английского языка, усвоить терминологию, характерную для использования в таких инновационных областях, как «наука о данных», «машинное обучение», «искусственный интеллект», «информационная безопасность», а также познакомиться с наиболее современными и набирающими популярность в сфере ИТ профессиями.

Работа над материалами учебно-практического пособия обеспечивает дальнейшее формирование коммуникативной компетенции, предполагающей развитие способности принимать участие в эффективном общении.

INTERNET RESOURCES

1. What is Data Science. − URL: https://www.analytixlabs.co.in/blog/what-is-data-science/ (дата обращения: 12.01.2025).

2. What is a Data Scientist? What Do They Do? − URL: https://www.techtarget.com/searchenterpriseai/definition/data-cientist (дата обращения: 12.01.2025).

3. Data scientist Job Description: Role, Responsibilities and More. – URL: https://www.simplilearn.com/data-scientist-job-description-article (дата обращения: 15.01.2025).

4. What is Data Profiling?| IBM - URL: https://www.ibm.com/think/topics/data-profiling (дата обращения: 15.01.2025).

5. What is Machine Learning: Definition, Types, Applications. − URL: https://www.analytixlabs.co.in/blog/what-is-machine-learning/ (дата обращения: 20.01.2025).

6. What is Artificial Intelligence (AI)?| Built in. – URL: https://builtin.com/artificial-intelligence (дата обращения: 20.01.2025)

7. What is Cybersecurity? [Definition, Importance, Types, Challenges]. – URL: https://www.techgeekbuzz.com/blog/what-is-cybersecurity/ (дата обращения: 26.01.2025).

8. Information Security Engineer. – URL: https://www.mygreatlearning.com/blog/information-security-engineer/ (дата обращения: 30.01.2025).

9. What are neural networks and what do we need them for? / Habr. – URL: https://habr.com/en/articles/676266/ (дата обращения: 08.02.2025).

10. A text neuro editor and other productivity tools have …| AKM EN. – URL: https://www.akm.ru/eng/press/a-text-neuro-editor-and-other-productivity-tools-have-appeared-in-yandex-browser/ (дата обращения: 15.02.2025).

11. 5 main challenges with datasets and AI ethics – Tooploox. – URL: https://tooploox.com/datasets-and-ai-ethics (дата обращения: 20.02.2025).

Учебное электронное издание

КОЙКОВА Татьяна Ивановна

СПЕЦИАЛЬНОСТИ В СФЕРЕ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

MAJORS IN IT

Учебно-практическое пособие

Издается в авторской редакции

Системные требования: Intel от 1,3 ГГц; Windows XP/7/8/10;
Adobe Reader; дисковод CD-ROM.

Тираж 9 экз.

Издательство Владимирского государственного университета
имени Александра Григорьевича и Николая Григорьевича Столетовых.
600000, Владимир, ул. Горького, 87.