



Big Data Analysis and Artificial Intelligence

Working program of the academic discipline (Syllabus)

Details of the academic discipline

Level of higher education	<i>Second (master)</i>
Branch of knowledge	<i>05 Social and behavioral sciences</i>
Specialty	<i>054 Sociology</i>
Educational program	<i>Social Data Analytics</i>
Discipline status (code)	<i>Compulsory, Professional Training Cycle</i>
Form of education	<i>full-time</i>
Year of training, semester	<i>2nd year, autumn semester</i>
Scope of the discipline	<i>4 ECTS credits, 120 hours</i> <i>Lectures: 18 hours, practical classes: 36 hours, independent work: 66 hours.</i>
Semester control/ control measures	<i>Final test, Modular Control Work</i>
Lessons schedule	<i>rozklad.kpi.ua</i> <i>1 hour of lectures and 2 hours of computer workshops per week</i>
Language of teaching	<i>English</i>
Information about head of the course / teachers	Lecturer: PhD, Associate Professor Ivan Pyshnograiev, pyshnograiev@wdc.org.ua Computer workshops: PhD, Associate Professor Ivan Pyshnograiev
Placement of the course	Google classroom https://classroom.google.com/c/NzA3MzMzMzNzY5NjE5

Program of educational discipline

1. Description of the educational discipline, its purpose, subject of study and learning outcomes

Discipline is normative in the educational program. The study of the academic discipline is aimed at the formation, development and consolidation of the acquirers of the following general and professional competencies:

ЗК 01 Ability to abstract thinking, analysis and synthesis,

ФК 02 Ability to identify, diagnose and interpret social problems of Ukrainian society and the world community,

ФК 04 Ability to collect and analyze empirical data using modern methods of sociological surveys,

ФК 11 Ability to perform open source data analysis (OSINT), analyze qualitative information, textual data, use data mining for social data,

ФК 12 Ability to apply modern statistical methods, models, digital technologies, specialized software for modeling social processes.

Upon completion of the course, the student should be able to demonstrate the following program learning outcome of the Educational and Scientific Program:

ПРП 04 Apply scientific knowledge, sociological and statistical methods, digital technologies, specialized software to solve complex problems of sociology and related fields of knowledge,

ПРП 05 Search, analyze and evaluate the necessary information in the scientific literature, data banks and other sources,

ПРП 12 Conduct open source data analysis (OSINT), analyze qualitative information, text data, use data mining for social data,

ПРП 14 Apply programming languages R and Python to analyze social data.

*At the end of the course, the student should **know**:*

- features of working with big data;
- methods of processing and analyzing big data;
- features of the application of artificial intelligence in the analysis of social data;
- methods of creating and applying machine learning models for data analysis.;

be able:

- to analyze big data using R and Python programming languages;
- to create models for analyzing social data.

2. Pre-requisites and post-requisites of the discipline (place in the structural and logical scheme of training according to the relevant educational program)

The discipline is based on the knowledge and skills of related disciplines studied in the previous semester and educational level. This discipline precedes ПО 12 "Execution of the master's thesis" can be one of its main components.

3. Content of the academic discipline:

Section 1. Introduction to big data analysis and artificial intelligence

Topic 1.1. Basic concepts of big data.

1. Characteristics and differences of big data;
2. Use of big data in Data Science;
3. Basic tools for working with big data;
4. Restrictions on the use of big data.

Topic 1.2. Basic concepts of artificial intelligence.

1. Basic definitions and areas of application;
2. Ethical problems when using artificial intelligence;
3. Basic tools for working with artificial intelligence models.

Section 2. The use of big data and artificial intelligence in social research

Topic 2.1. Computational methods in social research.

1. Main tasks and challenges;
2. Examples of application of machine learning and artificial intelligence in social research.

Topic 2.2. Analysis of social networks.

1. Statement of the problem and formalization of the network;
2. Methods and tools of social network analysis;
3. Analysis and forecasting of the development of social networks.

Topic 2.3. Agent modeling in social research.

1. Statement of the problem and basic definitions;
2. Modeling the behavior of a closed society;
3. Construction and research of an artificial society.

Topic 2.4. Stages of project creation during quantitative social research.

List of computer workshops:

1. Solving basic problems using big data and artificial intelligence.
2. Analysis of the social network to solve the problem.
3. Modeling the behavior of an artificial society.
4. Conducting social research using artificial intelligence.

4. Educational materials and resources

Basic:

1. Zgurovsky, M., . Zaychenko, Y. (2020). *Big Data: Conceptual Analysis and Applications*. Cham, Switzerland: Springer. <https://link.springer.com/book/10.1007/978-3-030-14298-8>
2. Олещенко, Л. М. Технології оброблення великих даних. Конспект лекцій [Електронний ресурс] : навчальний посібник / Л. М. Олещенко ; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 5,55 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2021. – 227 с. <https://ela.kpi.ua/handle/123456789/42206>

3. SpringerLink (Online service), Zomaya, A. Y., & Sakr, S. (2017). *Handbook of Big Data Technologies* (1st ed. 2017.). Cham: Springer International Publishing. <https://link.springer.com/book/10.1007/978-3-319-49340-4>
4. *Методи аналізу «великих даних» : методичні рекомендації з курсу «Прикладні задачі аналізу великих даних» / уклад. Н. М. Кізілова. – Харків : ХНУ імені В. Н. Каразіна, 2023. – 92 с. <https://ekhnuir.karazin.ua/items/6b5998f2-7e48-4dfb-8fa7-99aa3a03a150>*
5. *Методи та системи штучного інтелекту: Навчальний посібник для студентів/Уклад. : А.С. Савченко, О. О. Синельников. – К. : НАУ, 2017. – 190 с. https://er.nau.edu.ua/bitstream/NAU/40676/1/Методи%20та%20системи%20штучного%20інтелекту%20_Навч_посібн.pdf*

Additional:

6. *R програмування // Електронний ресурс. Режим доступу: <https://coderlessons.com/tutorials/mashinnoe-obuchenie/r-programmirovaniye/r-programmirovaniye>*
7. *Virtualization Technology // Електронний ресурс. Режим доступу: <https://www.sciencedirect.com/topics/computer-science/virtualization-technology>*
8. *Apache Hadoop // Електронний ресурс. Режим доступу: <http://hadoop.apache.org/>*
9. *Stuart Russel, Peter Norvig. Artificial Intelligence: A Modern Approach 4rd Edition, 2020, 1408 p.*

Educational content

5. Methods of mastering an educational discipline (educational component)

Lectures

Section 1. Introduction to Big Data Analysis and Artificial Intelligence

Lecture 1. The Big Data Phenomenon: Characteristics and Role in the Modern World

Definition of Big Data. The «5V» model (Volume, Velocity, Variety, Veracity, Value) and its expansion. Differences between Big Data and traditional statistical data. Data generation sources: social networks, IoT, transactional systems. Limitations of using Big Data: representativeness issues, noise, accessibility. The role of Data Science as a science of data processing and analysis methods.

Self-study assignments:

1. *Find and analyze a case study of successful Big Data usage in the commercial or public sector.*
2. *Create a comparative table «Traditional Statistics vs Big Data» based on the following criteria: volume, collection method, structuredness, general population.*

Lecture 2. Ecosystem and Tools of Big Data Technologies

Overview of Big Data architecture. Concept of distributed computing. Basic tools and frameworks (Hadoop, Spark – overview). Types of data storage: Data Warehouse and Data Lake. Cloud solutions for Big Data. Programming languages for data analysis (R, Python): advantages and disadvantages in the context of large datasets. Visualization as a tool for interpreting Big Data.

Self-study assignments:

1. *Review the documentation of Apache Spark or the Hadoop ecosystem, identify the main components.*
2. *Install a development environment (RStudio or Jupyter Notebook) and load a test dataset to check settings.*

Lecture 3. Artificial Intelligence: Concepts, History, and Ethics

Definition of Artificial Intelligence (AI). AI classification: Weak (Narrow AI), Strong (General AI), and Artificial Superintelligence. Main areas of application. Ethical challenges: algorithmic bias, decision-making transparency («black box»), data privacy, copyright. AI regulation (EU AI Act and other initiatives).

Self-study assignments:

1. *Analyze an ethical dilemma using a real-world scandal example (e.g., Cambridge Analytica or discrimination in hiring algorithms).*
2. *Familiarize yourself with the main provisions of the «Trustworthy AI» concept.*

Lecture 4. Artificial Intelligence Models and Tools

Overview of approaches to creating AI: logical approach, evolutionary methods, neural networks. Concept of Deep Learning. Large Language Models (LLM) and Generative AI: working principles and prompt engineering. Libraries for working with models (TensorFlow, PyTorch, scikit-learn – overview). Using APIs to integrate AI solutions.

Self-study assignments:

1. *Register on the Hugging Face or OpenAI platform, test the operation of one of the open models.*

2. Compile a list of tasks in your own scientific activity that can be automated using existing AI tools.

Section 2. Use of Big Data and Artificial Intelligence in Social Research

Lecture 5. Computational Methods in Social Research (Computational Social Science)

The paradigm of Computational Social Science. Main tasks: classification, clustering, regression in the context of social data. Machine Learning (ML) in sociology: supervised and unsupervised learning. Text analysis (NLP): sentiment analysis, topic modeling. Challenges regarding the validity of digitally obtained data.

Self-study assignments:

1. Find a scientific article where the NLP (Natural Language Processing) method is applied to analyze social moods.
2. Describe the difference between supervised and unsupervised learning using a sociological task example (e.g., electorate segmentation).

Lecture 6. Social Network Analysis (SNA): Formalization and Metrics

Graph theory as the basis of SNA. Basic network elements: nodes, edges, directionality, tie weight. Adjacency matrices. Levels of analysis: node level, group level, network level. Basic centrality measures: Degree, Closeness, Betweenness. Graph visualization.

Self-study assignments:

1. Construct an ego-network of your own environment (schematically or using software) and highlight key actors.

Lecture 7. Advanced Analysis and Forecasting of Social Networks

Community detection: algorithms and cluster interpretation. The «Small World» phenomenon and scale-free networks. Information spread and diffusion of innovations in networks. Link Prediction. Tools for SNA (Gephi, igrph, tidygraph).

Self-study assignments:

1. Download and install Gephi (or the corresponding package in R), familiarize yourself with the interface.
2. Find an example of a study on the spread of disinformation or viral content using network analysis methods.

Lecture 8. Agent-Based Modeling (ABM) of Social Processes

Concept of agent and environment. Principles of agent-based modeling: autonomy, interaction, bounded rationality. From micro-behavior to macro-phenomena (emergence). Classical models: Schelling's segregation model, "Predator-Prey" model in a social context. ABM tools (NetLogo, specialized libraries).

Self-study assignments:

1. Run a simulation of Schelling's segregation model, changing tolerance parameters, and describe the changes in results.
2. Propose an idea for an agent-based model describing crowd behavior or rumor spreading.

Lecture 9. Lifecycle of a Quantitative Social Research Project Using Big Data

Stages of a Data Science project: from hypothesis formulation to implementation. ETL processes in social research. Data Wrangling and handling missing values. Building analytical dashboards (Shiny, Tableau, PowerBI) to present results to stakeholders. Interpretation of machine learning results for social sciences.

Self-study assignments:

1. Develop a conceptual scheme (pipeline) of your own research project: determine the data source, processing methods, and expected result.
2. Familiarize yourself with examples of interactive dashboards displaying social or demographic indicators.

Practical classes (computer workshops)

Computer Workshop 1. Solving Basic Tasks Using Big Data and Artificial Intelligence. (6 hours)

Goal: To learn how to set up the environment for working with data, perform initial processing, and apply basic machine learning algorithms.

Plan

1. Preparation of the environment and data.
2. Data preprocessing (cleaning, descriptive statistics calculations, handling missing values).
3. Application of AI methods for solving the task.
4. Calculation of basic accuracy metrics. Interpretation of obtained results, coefficients, or feature importance.

Computer Workshop 2. Social Network Analysis for Solving a Defined Problem. (8 hours)

Goal: To master Social Network Analysis (SNA) methods for identifying influential agents and communities within a social structure.

Plan

1. *Problem formalization, data search, and loading.*
2. *Network construction and calculation of specialized metrics.*
3. *Detection of "communities" for solving the task.*
4. *Construction of network visualization with node coloring and sizing depending on the defined task.*

Computer Workshop 3. Modeling the Behavior of an Artificial Society. (8 hours)

Goal: To create and investigate an agent-based model to understand emergent social phenomena.

Plan

1. *Problem formulation. Definition of agents and their properties.*
2. *Model initialization with random agent distribution, observation of dynamics.*
3. *Conducting an experiment, analysis of the consequences of changing agent properties.*
4. *Analysis of the obtained final states.*

Computer Workshop 4. Conducting Social Research Using Artificial Intelligence. (10 hours)

Goal: To implement a full cycle of a mini-research project using AI methods for analyzing social processes.

Plan

1. *Problem formulation. Collection of unstructured data (texts) and structured data.*
2. *Data Mining. Building a conceptual model, selection of tools for conducting research.*
3. *Conducting research using machine learning models, LLMs, agent-based modeling, etc.*
4. *Formulation of sociological conclusions based on obtained patterns.*

Modular Control Work (2 hours)

Final test (2 hours)

6. The student's independent work

The student's independent work includes:

- *preparation for classroom sessions – 62 hours;*
- *preparation for the modular control work – 2 hours.*

Total – 66 hours.

Policy and control

7. Policy of academic discipline (educational component)

Students must attach all assignments in their personal Google Classroom account. Deadlines for each assignment are indicated in the tasks within Google Classroom. Assignments must be completed in compliance with academic integrity. The policy and principles of academic integrity, and the ethical behavior of students, are defined in the Honor Code <https://kpi.ua/code>. The lecturer may suggest that students take online courses on the Coursera platform. Additionally, certificates from these courses may be partially credited in accordance with the Regulations https://document.kpi.ua/files/2020_7-124.pdf.

The topics of the assignments are aimed at deepening the understanding of the lecture material. During computer workshops, problems and exercises related to the lecture topics are solved.

8. Types of control and rating system for evaluating learning outcomes (ELO)

Semester control: *Final test.*

The student's semester rating *for the discipline is assigned by the lecturer and consists of points received for:*

- ~ *completion of the modular control work;*
- ~ *completion of 4 computer workshops.*

Criteria for awarding points during the semester:

1. *The modular control work is valued at 20 points.*
2. *Each of the workshops is valued at 20 points.*

Criteria for awarding points for control measures:

- **«Excellent»: 95-100%** – the student demonstrated comprehensive, systematic, and deep knowledge of the educational material of the discipline; demonstrated the ability to freely perform all tasks provided by the program; mastered primary and supplementary literature; showed creative abilities in understanding, and in the logical, clear, concise, and distinct interpretation of the educational material; mastered the relationship of the basic concepts of the discipline and their significance for further professional activity.
- **«Very good»: 85-94%** – the student demonstrated systematic knowledge of the educational material of the discipline above the average level; demonstrated the ability to perform all tasks provided by the program well, while making minor errors; mastered primary and supplementary literature; mastered the relationship of the basic concepts of the discipline and their significance for further professional activity.
- **«Good»: 75-84%** – the student demonstrated generally good knowledge of the educational material when performing the tasks provided by the program, but made a number of noticeable errors; mastered primary literature; showed a systematic character of knowledge in the discipline; is capable of their independent use and replenishment in the process of further educational work and professional activity.
- **«Satisfactory»: 65-74%** – the student demonstrated knowledge of the main educational material of the discipline in the volume necessary for further study and future professional activity; familiarized themselves with the primary literature; coped with the completion of tasks provided by the program, but made a significant number of errors or shortcomings in answers to questions during interviews, testing, and task execution, the principal ones of which they can eliminate independently.
- **«Sufficient»: 60-64%** – the student demonstrated knowledge of the main educational material of the discipline in the minimum volume necessary for further study and future professional activity; familiarized themselves with the primary literature; mainly completed the tasks provided by the program, but made errors in answers to questions during interviews, testing, and task execution, which they can eliminate only under the guidance and with the help of the teacher.
- **«Unsatisfactory»: 30-59%** – the student had significant gaps in the knowledge of the main educational material; made principal errors when performing the tasks provided by the program, but is capable of independently reworking the program material and preparing for the retake of the discipline.
- **«Unsatisfactory»: 0-29%** – the student had no knowledge of a significant part of the educational material of the discipline; made principal errors when performing the majority of the tasks provided by the program or did not perform these tasks.

The condition for the first attestation is a current rating of at least 30% of the planned points for the semester.

The condition for the second attestation is a current rating of at least 50% of the planned points.

A necessary condition for admission to the Final test is the successful completion (acceptance) of all computer workshops and a semester rating of 60 points. In this case, the Final test is awarded based on the semester rating result.

If the student has not obtained 60 points or wishes to improve their grade, it is possible to write a **final test**. The credit work is valued at **100 points**. The credit is conducted in the form of a written work containing three theoretical questions and two practical ones. Each task is valued at 20 points according to the following criteria:

- **«Excellent»**, complete answer, at least 90% of the required information, executed according to the «skills» level requirements (complete, error-free solution of the task) – **18-20 points**;
- **«Good»**, sufficiently complete answer, at least 75% of the required information, executed according to the «skills» level requirements or containing minor inaccuracies (complete solution of the task with minor inaccuracies) – **14-16 points**;
- **«Satisfactory»**, incomplete answer, at least 60% of the required information, executed according to the «stereotypical» level requirements and containing some errors (task executed with certain shortcomings) – **12 points**;
- **«Unsatisfactory»**, the answer does not meet the conditions for «satisfactory» – **0 points**.

The sum of rating points obtained by the student during the semester is converted into a final grade according to the table.

Table of correspondence of rating points to grades on the university scale:

Бали:	Оцінка
100...95	Excellent
94...85	Very good

<i>84...75</i>	<i>Good</i>
<i>74...65</i>	<i>Satisfactorily</i>
<i>64...60</i>	<i>Enough</i>
<i>Less than 60</i>	<i>Unsatisfactorily</i>
<i>Failure to fulfill the conditions of admission to the semester control</i>	<i>Not allowed</i>

Working program of the academic discipline (syllabus):

Compiled by Ivan Pyshnograiev, Associate Professor, Candidate of Physical and Mathematical Sciences.



Approved by the Department of AI (protocol № 14 from 11.05.2024)

Agreed by the Methodical Commission of the ES IASA (protocol № 10 from 24.06.2024)