Національний технічний університет України «КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ імені ІГОРЯ СІКОРСЬКОГО»

КАФЕДРА ШТУЧНОГО ІНТЕЛЕКТУ

Department of Artificial Intelligence

BIG DATA ANALYSIS AND ARTIFICIAL INTELLIGENCE

Curriculum (Syllabus)

Course details

Level of higher education	***Second (Master's)***
Field of knowledge	*C - social sciences, journalism, information and international relations*
Specialisation	*C5 Sociology*
Educational programme	*Social Data Analytics*
Status of discipline (code)	*Mandatory, professional training cycle*
Form of study	*Full-time (day)*
Year of training, semester	*2nd year, autumn semester*
Scope of the discipline	*4 ECTS credits, 120 hours* *16 hours of lectures, 30 hours of practical classes (computer practice), 74 hours of independent work.*
Semester assessment/assessment measures	*Test, Modular control work*
Class schedule	*https://schedule.kpi.ua/* *1 hour of lectures and 2 hours of computer practicals per week*
Language of instruction	*Ukrainian*
Information about course coordinator/teachers	Lecturer: *Ivan Oleksandrovych Pyshnohraiev, PhD in Physics and Mathematics, Associate Professor, pyshnograiev@wdc.org.ua* Computer workshops: *Ivan Oleksandrovych Pyshnohraiev, PhD in Physics and Mathematics, Associate Professor*
Course location	Google Classroom https://classroom.google.com/c/NzA3MzMzNzY5NjE5

Course programme

1. Description of the course, its purpose, subject matter and learning outcomes

The discipline is a compulsory part of the educational programme. The study of the discipline is aimed at forming, developing and consolidating the following general and professional competences in students:
GC 01 Ability to think abstractly, analyse and synthesise
FC 02 Ability to identify, diagnose and interpret social problems of Ukrainian society and the world community,
PC 04 Ability to collect and analyse empirical data using modern methods of sociological research,
FC 11 Ability to analyse open source data (OSINT), analyse qualitative information, text data, use intellectual analysis for social data,
FK 12 Ability to apply modern statistical methods, models, digital technologies, and specialised software for modelling social processes.

As a result of studying the course, students should be able to demonstrate the following learning outcomes:
PRN 04 Apply scientific knowledge, sociological and statistical methods, digital technologies, specialised software to solve complex problems in sociology and related fields of knowledge,
PRN 05 Search for, analyse and evaluate necessary information in scientific literature, databases and other sources.

PRN 12 Analyse open source data (OSINT), analyse qualitative information, text data, use intelligent analysis for social data,
PRN 14 Apply the R and Python programming languages to analyse social data.

*At the end of the course, students should **know**:*
- *the specifics of working with big data;*
- *methods of processing and analysing big data;*
- *features of applying artificial intelligence in social data analysis;*
- *methods of creating and applying machine learning models for data analysis.*

***be able to**:*
- *analyse big data using the R and Python programming languages;*
- *create models for social data analysis.*

2. Prerequisites and post-requisites of the discipline (place in the structural-logical scheme of training under the relevant educational programme)

The discipline is based on the knowledge and skills of related disciplines studied in the previous year and educational level. This discipline precedes OK PO 12 "Master's Thesis" and may be one of its main components.

3. Contents 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. *Limitations of big data use.*

Topic 1.2. Basic concepts of artificial intelligence.
1. *Basic definitions and areas of application;*
2. *Ethical issues in the use of artificial intelligence;*
3. *Basic tools for working with artificial intelligence models.*

Section 2. 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 the application of machine learning and artificial intelligence in social research.*

Topic 2.2. Social network analysis.
1. *Problem formulation and network formalisation;*
2. *Methods and tools for analysing social networks;*
3. *Analysis and forecasting of social network development.*

Topic 2.3. Agent-based modelling in social research.
1. *Problem formulation and basic definitions;*
2. *Modelling the behaviour of a closed society;*
3. *Construction and study 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 social networks to solve a given problem.*
3. *Modelling the behaviour of an artificial society.*
4. *Conducting social research using artificial intelligence.*

4. Teaching 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. *Oleshchenko, L. M. Big Data Processing Technologies. Lecture Notes [Electronic resource]: textbook / L. M. Oleshchenko; Igor Sikorsky KPI. – Electronic text data (1 file: 5.55 MB). – Kyiv: Igor Sikorsky Kyiv Polytechnic Institute, 2021. – 227 p. 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. *Methods of big data analysis: methodological recommendations from the course "Applied problems of big data analysis" / compiled by N. M. Kizilova. – Kharkiv: V. N. Karazin Kharkiv National University, 2023. – 92 p. https://ekhnuir.karazin.ua/items/6b5998f2-7e48-4dfb-8fa7-99aa3a03a150*
5. *Methods and Systems of Artificial Intelligence: A Textbook for Students / Compiled by: A. S. Savchenko, O. O. Sinelnikov. – Kyiv: NAU, 2017. – 190 p. https://er.nau.edu.ua/bitstream/NAU/40676/1/Методи%20та%20системи%20штучного%20інтелекту%20_Навч_посібн.pdf*

Supplementary:

6. *R programming // Electronic resource. Access mode: https://coderlessons.com/tutorials/mashinnoe-obuchenie/r-programmirovanie/r- programmirovanie*
7. *Virtualisation Technology // Electronic resource. Access mode: https://www.sciencedirect.com/topics/computer-science/virtualization-technology*
8. *Apache Hadoop // Electronic resource. Access mode: http://hadoop.apache.org/*
9. *Stuart Russel, Peter Norvig. Artificial Intelligence: A Modern Approach 4rd Edition, 2020, 1408 p.*

Educational content

5. Methodology for mastering the academic discipline (educational component)

Lectures

Section 1. Introduction to big data analysis and artificial intelligence

Lecture 1. The phenomenon of big data: characteristics and role in the modern world. Definition of the concept of Big Data. The "5V" model (Volume, Velocity, Variety, Veracity, Value) and its extension. The difference between big data and traditional statistical data. Sources of data generation: social networks, IoT, transaction systems. Limitations of big data use: the problem of representativeness, noise, accessibility. The role of Data Science as a science of data processing and analysis methods.

Assignment for independent study:

1. *Find and analyse a case study of the successful use of Big Data in the commercial or public sector.*
2. *Compile a comparative table "Traditional Statistics vs Big Data" based on the following criteria: volume, collection method, structure, general population.*

Lecture 2. *Big Data ecosystem and tools*

Overview of the architecture for working with big data. The 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 arrays. Visualisation as a tool for interpreting big data.

Assignment for independent study:

1. *Familiarise yourself with the Apache Spark or Hadoop ecosystem documentation and identify the main components.*
2. *Install a development environment (RStudio or Jupyter Notebook) and download a test dataset to check the settings.*

Lecture 3. *Artificial intelligence: concepts, history and ethics*

Definition of artificial intelligence (AI). Classification of AI: narrow AI, general AI, and artificial superintelligence. Main areas of application. Ethical challenges: algorithm bias, transparency of decision-making ("black box"), data privacy, copyright. AI regulation (EU AI Act and other initiatives).

Assignment for independent study:

1. *Analyse an ethical dilemma using a real-life scandal as an example (e.g. Cambridge Analytica or discrimination in hiring algorithms).*
2. *Familiarise yourself with the main provisions of the Trustworthy AI concept.*

Lecture 4. *Artificial intelligence models and tools*

Overview of approaches to AI creation: logical approach, evolutionary methods, neural networks. The concept of deep learning. Large language models (LLM) and generative AI: principles of operation and prompt engineering. Libraries for working with models (TensorFlow, PyTorch, scikit-learn – overview). Using APIs to integrate AI solutions.

Homework assignment:

1. *Register on the Hugging Face or OpenAI platform, test one of the open models.*
2. *Make a list of tasks in your own scientific activity that can be automated using existing AI tools.*

Section 2. Using big data and artificial intelligence in social research

Lecture 5. *Computational methods in social research (Computational Social Science)*

The paradigm of computational social sciences. Main tasks: classification, clustering, regression in the context of social data. Machine learning in sociology: supervised and unsupervised learning. Text analysis (NLP): sentiment, topic modelling. Challenges of the validity of digitally obtained data.

Assignment for the course:

1. *Find a scientific article that uses NLP (natural language processing) to analyse social attitudes.*
2. *Describe the difference between supervised and unsupervised learning using a sociological task as an example (e.g., voter segmentation).*

Lecture 6. *Social network analysis (SNA): formalisation and metrics*

Graph theory as the basis of SNA. Basic network elements: nodes, edges, directionality, link weight. Adjacency matrices. Levels of analysis: node level, group level, network level. Basic centrality measures: degree, closeness, betweenness. Graph visualisation.

Homework assignment:

1. *Build an ego-network of your own environment (schematically or using software) and identify key actors.*

Lecture 7. *In-depth analysis and forecasting of social networks*

Community detection: algorithms and interpretation of clusters. The "small world" phenomenon and scale-free networks. Information dissemination and diffusion of innovations in networks. Link prediction. Tools for SNA (Gephi, igraph, tidygraph).

Homework assignment:

1. *Download and install Gephi (or the corresponding package in R), familiarise yourself with the interface.*
2. *Find an example of research on the spread of misinformation or viral content using network analysis methods.*

Lecture 8. *Agent-based modelling (ABM) of social processes. The life cycle of a quantitative social research project using Big Data*

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

Stages of a data science project: from hypothesis formulation to implementation. ETL processes (Extract, Transform, Load) in social research. Data cleaning (data wrangling) and working with gaps. Building analytical dashboards

(Shiny, Tableau, PowerBI) to present results to stakeholders. Interpretation of machine learning results for social sciences.

Assignment for the course:

1. *Run a simulation of Schelling's segregation model, changing the tolerance parameters, and describe the changes in the results.*
2. *Propose an idea for an agent-based model that describes crowd behaviour or the spread of rumours.*
3. *Develop a conceptual pipeline for your own research project: identify the data source, processing methods, and expected results.*
4. *Familiarise yourself with examples of interactive dashboards that display social or demographic indicators.*

Practical classes (computer workshops)

Computer workshop 1. *Solving basic problems using big data and artificial intelligence. (6 hours)*

Objective: *Learn how to set up an environment for working with data, perform initial processing, and apply basic machine learning algorithms.*

Plan

1. Preparing the environment and data.

2. Preliminary data processing (cleaning, descriptive calculations, cleaning gaps).

3. Application of AI methods to solve the problem.

4. Calculation of basic accuracy metrics. Interpretation of the results, coefficients, or feature importance.

Computer workshop 2. *Analysis of social networks to solve the problem. (6 hours)*

Objective: *To master social network analysis (SNA) methods for identifying influential agents and communities in the social structure.*

Plan

1. Formalisation of the task, search and data download.

2. Network construction and calculation of specialised metrics.

3. Identification of "communities" to solve the problem.

4. Building a network visualisation with node colouring and size depending on the task at hand.

Computer workshop 3. *Modelling the behaviour of an artificial society. (8 hours)*

Objective: *To create and explore an agent-based model for understanding emergent social phenomena.*

Plan

1. Setting the task. Defining agents and their properties.

2. Initialisation of the model with random distribution of agents, observation of dynamics.

3. Conducting an experiment, analysing the consequences of changes in agent properties.

4. Analysis of the final states obtained.

Computer workshop 4. *Conducting social research using artificial intelligence (8 hours).*

Objective: *To implement a full cycle of mini-research using AI methods to analyse social processes.*

Plan

1. Setting the task. Collecting unstructured data (texts) and structured data.

2. Data mining. Building a conceptual model, selecting tools for conducting research.

3. Conducting research using machine learning models, LLM, agent modelling, etc.

4. Formulating sociological conclusions based on the patterns obtained.

Modular control work 2 hours.

6. Independent work by the student

Independent work by the applicant includes:

- preparation for classroom sessions – 62 hours;

- preparation for the Modular control work – 4 hours.

Total – 66 hours.

Policy and control

7. Academic discipline policy (educational component)

All work must be uploaded by students to their personal Google Classroom account. The deadlines for each assignment are indicated in the assignments in Google Classroom. Work must be completed in accordance with academic integrity. The policy and principles of academic integrity and ethical behaviour of students are defined in the Code of Honour https://kpi.ua/code. At the request of the applicant, in conditions that do not facilitate regular attendance, it is permissible to study individual content-rich parts of the discipline in asynchronous mode, in particular through distance learning courses and other forms of informal learning. In order for the credits for such courses to be taken into account in the rating system, they must correspond in content to certain topics of the syllabus, and their completion must be agreed with the teacher of the discipline. To confirm completion of informal learning, the student must provide a relevant document (certificate) indicating the name of the courses and their duration in hours. Recognition of the results of informal education takes place in accordance with the procedure set out in the relevant Regulations of Igor Sikorsky KPI: https://osvita.kpi.ua/node/179.

The topics of the assignments are aimed at deepening the material covered in the lectures. During the computer practical classes, problems and exercises on the topics of the lectures are solved.

8. Types of control and rating system for assessing learning outcomes (RSO)

Semester assessment: ***test****.*

The student's semester rating for the discipline is given by the lecturer and consists of points awarded for:

~ completion of the modular control work;

~ completion of 4 computer workshops.

Criteria for awarding points for the semester:

1) The Modular control work is worth 20 points.

2) Each practical assignment is worth 20 points.

Criteria for awarding points for tests:

- *"Excellent": 95-100% - the student has demonstrated comprehensive, systematic and in-depth knowledge of the subject matter; demonstrated the ability to freely perform all tasks required by the programme; mastered the main and additional literature; showed creative abilities in understanding, logical, clear, concise and clear interpretation of the course material; mastered the interconnection of the main concepts of the discipline, their significance for further professional activity*
- *"very good": 85-94% - the applicant has demonstrated systematic knowledge of the subject matter above average; demonstrated the ability to perform all tasks required by the programme well, with minor errors; mastered the main and additional literature; mastered the interrelationship of the main concepts of the discipline and their significance for further professional activity*
- *"Good": 75-84% - the applicant demonstrated generally good knowledge of the course material when performing the tasks provided for in the programme, but made a number of noticeable mistakes; mastered the main literature; demonstrated systematic knowledge of the discipline; is capable of using and supplementing this knowledge independently in the process of further study and professional activity*
- *"Satisfactory": 65-74% - the applicant demonstrated knowledge of the basic course material to the extent necessary for further study and future professional activity; familiarised themselves with the main literature; has coped with the tasks set out in the programme, but has made a significant number of mistakes or shortcomings in questions during interviews, tests and tasks, etc., the fundamental ones of which can be eliminated independently*
- *"sufficient": 60-64% - the applicant demonstrated knowledge of the basic educational material of the discipline to the minimum extent necessary for further study and future professional activity; familiarised themselves with the basic literature; has basically completed the tasks specified in the programme, but has made mistakes in answering questions during interviews, testing and performing tasks, etc., which he can only correct under the guidance and with the help of a teacher*

- *"unsatisfactory": 30-54% - the applicant had significant gaps in knowledge of the basic course material; made fundamental mistakes in completing the tasks specified in the programme, but is able to independently complete the programme material and prepare to retake the course*
- *"unsatisfactory": 0-29% - the applicant did not have knowledge of a significant part of the course material; made fundamental mistakes in performing most of the tasks specified in the programme or did not perform these tasks*

The condition for the first assessment is a current rating of at least 30% of the planned points for the semester. The condition for the second assessment is a current rating of at least 50% of the planned points.

A necessary condition for admission to the exam is the completion of all computer practicals and a semester rating of 60 points. In this case, the exam is graded based on the semester rating.

If the applicant has not scored 60 points or wants to improve their grade, it is possible to write a credit paper. The credit paper is scored out of 100 points. The credit is conducted in the form of a written paper, which includes three theoretical questions and two practical ones. Each task is scored out of 20 points according to the following criteria:

- *"excellent", complete answer, at least 90% of the required information, performed in accordance with the requirements for the "skills" level (complete, error-free solution of the task) – 18-20 points;*
- *"good", sufficiently complete answer, at least 75% of the required information, performed in accordance with the requirements for the "skills" level, or there are minor inaccuracies (complete solution of the task with minor inaccuracies) – 14-16 points;*
- *"satisfactory", incomplete answer, at least 60% of the required information, performed in accordance with the requirements for the "stereotypical" level and some errors (task performed with certain shortcomings) – 12 points;*
- *"unsatisfactory", the answer does not meet the requirements for "satisfactory" – 0 points.*

The sum of the rating points received by applicants during the semester is converted to a final grade according to the table.

Table of correspondence between rating points and grades on the university scale:

Points:	***Grade***
100…95	*Excellent*
94	*Very good*
84	*Good*
74…65	*Satisfactory*
64	*Sufficient*
Less than 60	*Unsatisfactory*
Failure to meet the conditions for admission to semester control	*Not admitted*

Work programme for the academic discipline (syllabus):

Compiled by *Associate Professor, Candidate of Physical and Mathematical Sciences, Associate Professor Ivan Oleksandrovych Pyshnohraiev*

Approved by *the AI Department (Minutes No. 14 of 24 June 2025)*

Approved by *the Methodological Commission of the IPSA (Minutes No. 7 of 25 June 2025)*