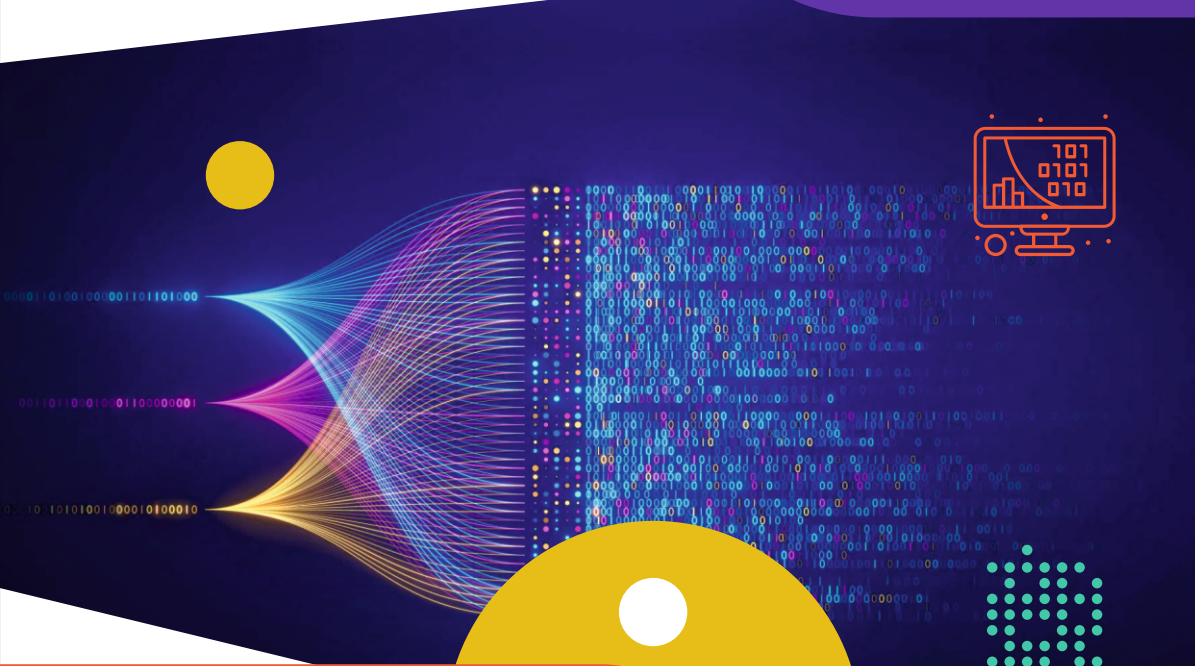




Shakhmardan  
Yessenov  
Foundation



Yessenov  
Data Lab



16 June -  
11 July

Almaty 2025



Shakhmardan  
Yessenov  
Foundation

## Table of contents

- What is Yessenov Data Lab 2025 Summer School?
- Training program

**Week 1** Python and Chatbots

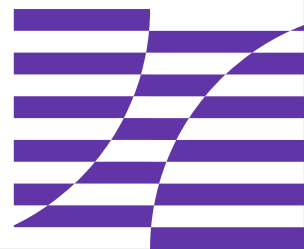
**Week 2** Machine Learning

**Week 3** Large Language  
Models (LLM)

**Week 4** Computer Vision



Yessenov  
Data Lab





Yessenov  
Data Lab



## What is Yessenov Data Lab 2025?

Yessenov Data Lab 2025 summer school is

- an annual summer school on artificial intelligence, where professionals from different countries teach best practices in working with big data and machine learning
- is a 4-week intensive. **2 weeks online training, 2 weeks offline training (in Almaty)**

**Dates of the event: 16 June – 11 July 2025**

**29 June – Teambuilding. Networking**

### DURING THE 4 WEEKS OF SCHOOL YOU WILL BE ABLE TO:

- learn big data analysis skills that you can apply in practice;
- communicate with lecturers, ask questions of interest;
- meet like-minded people and make new connections.

Participants will receive certificates from YDL school, as well as an academic transcript from AlmaU corresponding to international educational standards of the Bologna system (recognition in universities)

**This year we are focusing on NLP and Large Language Models (LLM)**



Shakhmardan  
Yessenov  
Foundation



Yessenov  
Data Lab

Online



## Training Program

Lecturer:

**Timur Bakibayev**

Ph.D. in Computer Science,  
Director of the Digital  
Transformation Institute,  
Narxoz



### Week 1

### Python and Chat Bots

16-20  
June

**Objective:** to learn the best practices of the Python programming language for working with data

#### Day 1 Programming Fundamentals for Data Science

- 10:00 – 10:10 Opening of the Summer School, welcome message
- 10:10 – 11:30 YDL Program Overview. Fundamentals of Artificial Intelligence
- 11:45 – 13:15 ChatGPT Principles
- 14:30 – 16:00 Laboratory Work with ChatGPT API
- 16:15 – 18:00 Laboratory Work Analysis

#### Day 2 Working with Data

- 10:00 – 11:30 NumPy and Pandas Libraries Review
- 11:45 – 13:15 Data Grouping. Filters, Sorting
- 14:30 – 16:00 Laboratory Work: Work with Pandas.
- 16:15 – 18:00 Laboratory Work Analysis

#### Day 3 Data Visualization

- 10:00 – 11:30 Matplotlib Library Review
- 11:45 – 13:15 Seaborn Library Review
- 14:30 – 16:00 Laboratory Work: Visualization
- 16:15 – 18:00 Laboratory Work Analysis

#### Day 4 Forecasting Basics

- 10:00 – 11:30 Regression Problem Review
- 11:45 – 13:15 Feature Importance in Model Building
- 14:30 – 16:00 Laboratory Work: Analysis and Data Preprocessing
- 16:00 – 17:00 Laboratory Work Analysis
- 17:00 – 18:00 Motivational Meeting

#### Day 5 Chat bots

- 10:00 – 11:30 Streamlit Review
- 11:45 – 13:15 How Chat bots work
- 14:30 – 16:00 Laboratory Work: Streamlit and Chat bots
- 16:00 – 18:00 Streamlit Deployment on the server (Deployment)



Shakhmardan  
Yessenov  
Foundation



Yessenov  
Data Lab

Online



## Training Program

Coach:

**Kuanysh Abeshev**

Ph.D. in Mathematics,  
Dean of the School of  
Digital Technologies, AlmaU



**Week 2**

**Machine Learning**

**23-27  
June**

**Objective:** to master the methods of predictive analysis and machine learning

### Day 1 Statistical Analysis

- 10:00 – 11:30 Exploratory analysis, statistics
- 11:45 – 13:15 Model evaluation metrics
- 14:30 – 16:00 Laboratory Work
- 16:15 – 18:00 Laboratory Work, discussion

### Day 2 Regression modelling

- 10:00 – 11:30 Linear Regression Basics. Metrics. Data Verification.
- 11:45 – 13:15 Data Preprocessing
- 14:30 – 16:00 Laboratory Work
- 16:15 – 18:00 Laboratory Work, discussion

### Day 3 Evaluating Performance - Classification

- 10:00 – 11:30 Metrics for Classification Problems
- 11:45 – 13:15 Logistic regression (Binary, Multi-Class)
- 14:30 – 16:00 Laboratory Work
- 16:00 – 18:00 Laboratory Works Analysis, Distribution of Datasets for the Project Defence

### Day 4 Unsupervised learning

- 10:00 – 11:30 Feature Importance
- 11:45 – 13:15 Unsupervised learning. Dimensionality Reduction
- 14:30 – 16:00 Laboratory Work
- 16:00 – 17:00 Laboratory Work, discussion
- 17:00 – 18:00 Motivational Meeting

### Day 5 Project on Regression Modelling

- 10:00 – 11:30 Complete Machine Learning Project
- 11:45 – 13:15 Q&A
- 14:30 – 16:00 Project Defence
- 16:00 – 18:00 Project Defence and Discussion of the Results

**29 June (Sunday). Almaty**

**10:00 – 14:00** Team building



Shakhmardan  
Yessenov  
Foundation



Yessenov  
Data Lab



Алматы

Offline

## Training Program

Lecturer:

**Nikita Saxena**

Google Deepmind Research  
Engineer



**Week 3**

**Large Language Models (LLM)**

**30 June -  
4 July**

**Objective:** theoretical and practical basics  
of large language models

### Day 1 Introduction to LLM

- 10:00 – 11:30 Transformers and the Self-Attention Mechanism
- 11:45 – 13:15 Sampling Strategies and Prompt Engineering
- 14:30 – 16:00 Laboratory Work
- 16:15 – 18:00 Laboratory Works Analysis

### Day 2 Pre- and post-training

- 10:00 – 11:30 Pre-training
- 11:45 – 13:15 Post-training (SFT and LoRA)
- 14:30 – 16:00 Laboratory Work
- 16:15 – 18:00 Laboratory Works Analysis

### Day 3 Alignment by preference

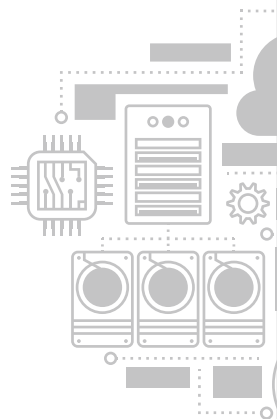
- 10:00 – 11:30 Introduction to RL and PPO
- 11:45 – 13:15 Reinforcement learning with feedback
- 14:30 – 16:00 Laboratory Work
- 16:00 – 18:00 Laboratory Works Analysis

### Day 4 Assessment and security

- 10:00 – 11:30 LLM assessment
- 11:45 – 13:15 Ethics and safety of AI
- 14:30 – 16:00 Laboratory Work
- 16:00 – 17:00 Laboratory Works Analysis
- 17:00 – 18:00 Motivational Meeting

### Day 5 Designing an application based on LLM

- 10:00 – 11:30 Introduction to Langchain
- 11:45 – 13:15 Laboratory Work
- 14:30 – 16:00 Project Implementation
- 16:00 – 18:00 Presentation of the Project



Shakhmardan  
Yessenov  
Foundation



Yessenov  
Data Lab



Алматы

Offline

## Training Program

Lecturer:

**Vladislav Yushchenko**

Senior ML Engineer,  
Meta London



### Week 4

### Computer Vision

7 -  
11 July

**Objective:** to master deep learning methods and neural networks

#### Day 1 Deep learning (MLP, ANN)

- 10:00 – 11:30 Introduction to Neural Networks. Multilayer Perceptron
- 11:45 – 13:15 Backpropagation Algorithm
- 14:30 – 16:00 Laboratory Work
- 16:15 – 18:00 Laboratory Work, discussion

#### Day 2 Deep Learning with PyTorch

- 10:00 – 11:30 Introduction to PyTorch Library
- 11:45 – 13:15 Convolutional Neural Networks (CNN). Image Classification. Metrics
- 14:30 – 16:00 Laboratory Work
- 16:15 – 18:00 Laboratory Work, discussion

#### Day 3 Segmentation

- 10:00 – 11:30 Data Preparation
- 11:45 – 13:15 Model Building, Training, Quality Metrics
- 14:30 – 16:00 Laboratory Work
- 16:00 – 18:00 Laboratory Work, discussion

#### Day 4 Object Detecion

- 10:00 – 11:30 Data Preparation
- 11:45 – 13:15 Model Building, Training, Quality Metrics
- 14:30 – 16:00 Laboratory Work
- 16:00 – 18:00 Laboratory Work, discussion

#### Day 5 Introduction to MLOps

- 10:00 – 11:30 Introduction to MLOps, Levels of Automation
- 11:45 – 13:15 Experiment Tracking and Model Deployment. MLOps Best Practices
- 14:30 – 16:00 Laboratory Work
- 16:00 – 17:00 Laboratory Work, discussion
- 17:00 - 18:00 Motivational Meeting. Closing

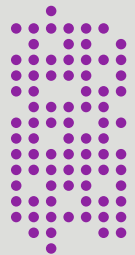


Shakhmardan  
Yessenov  
Foundation



Yessenov  
Data Lab

## Stay In Touch With Us



Shakhmardan  
Yessenov  
Foundation

STAY IN  
TOUCH  
WITH US

[yessenovfoundation.org](http://yessenovfoundation.org)  
[info@yessenovfoundation.org](mailto:info@yessenovfoundation.org)  
+7 771 759 5944