

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Суворов Сергей Дмитриевич
Должность: Ректор
Дата подписания: 31.08.2026 16:35:32
Уникальный программный ключ:
a39bdb15d680d3b0adbfced0af5c1efb14747dc0

Course Syllabus

Skoltech

Contact Person

Alexander Panchenko

Contact Person's E-mail

a.panchenko@skoltech.ru

Course Title (in English)

Transformers and Large Language Models

Course Title (in Russian)

Трансформерные архитектуры и большие языковые модели

Lead Instructor

Panchenko, Alexander

1. Annotation

Course Description

The course is about modern models for natural language processing. The new generation of neural network-based methods based on deep learning has dramatically improved the performance of a wide range of natural language processing tasks, ranging from text classification to question answering. The course covers the basics and the details of successful models and methods for natural language processing based on neural networks. Only limited coverage of non-Transformer based architectures will be given, such as word embeddings (which are subject to Introduction to NLP course). On the contrary, this course is focused on more sophisticated language models, such as BERT. Special attention is given to models based on the Transformer architecture, such as GPT, T5, BART, and the most recent LLMs. Typically we read a subset of lectures from this course: <https://s-nlp.github.io/transformers-course/>

Аннотация

Курс посвящен современным моделям обработки естественного языка. Новое поколение методов на основе нейронных сетей, основанных на глубоком обучении, значительно улучшило производительность широкого спектра задач обработки естественного языка, начиная от классификации текста и заканчивая ответами на вопросы. Курс охватывает основы и детали успешных моделей и методов обработки естественного языка на основе нейронных сетей. Будет дан лишь ограниченный обзор нетрансформаторных архитектур, таких как векторные представления слов (которые детально рассматриваются в рамках курса "Введение в обработку естественного языка"). Данный курс ориентирован на более современные языковые модели, такие как BERT. Особое внимание уделяется моделям, основанным на архитектуре Transformer, таким как GPT, T5, BART и самым последним LLMs. Как правило, курс представляет собой подмножество лекций из <https://s-nlp.github.io/transformers-course/>

2. Basic Information

Course Academic Level

BSc

MSc

PhD

Course Academic Level

Master-level course suitable for PhD students

Number of ECTS credits

3

Course Prerequisites / Recommendations

Machine learning, Python

Type of Assessment

Graded

Grading Scale

A:	90
B:	80
C:	70
D:	60
E:	50

Course Term (in context of Academic Year)

Term 4

Students of Which Programs do You Recommend to Consider this Course as an Elective?

BSc Programs	Masters Programs	PhD Programs
Math with AI	AI and Financial Technologies AI Robotics Advanced Computational Science Advanced Manufacturing Technologies Applied Computational Mechanics Data Science Engineering Systems Internet of Things and Wireless Technologies Life Sciences Materials Science Next Generation Communication Technologies Technological Entrepreneurship and Innovative Business Development	Computational and Data Science and Engineering Engineering Systems Life Sciences Materials Science and Engineering Mathematics and Mechanics

Maximum Number of Students

	Maximum Number of Students
Overall:	75
Per Group (for seminars and labs):	5

3. Course Content

Course contains Industry-oriented elements?

No

Invited lecturers from the Industry

ECTS Credit System – Reference Tool

Total Number of ast. Hours in Your Course81

Topic	Summary of Topic	Contact Hours: Lectures	Contact Hours: Seminars	Contact Hours: Labs	Non-contact Hours: Student's Independent Study
Intro	Introduction. Course logistics. Questions and answers.	2	2	2	2
Transformer	Transformer: motivation, the original architecture and attention mechanism.	2	2	2	2
Encoders	Transformer-based Encoders. Masked language models based on Transformer. BERT and related models	2	2	2	2
Classification	Classification and sequence tagging with Transformers. Using encoders to generate feature representation for various NLU tasks.	2	2	2	2
Decoders	Transformer-based Decoders. Transformer-based Decoders. Generation of text using Transformers. GPT and related decoders. Prompt tuning and beyond*	2	2	2	2
seq2seq	Text-to-text Transformers. Sequence to sequence tasks: machine translation, text detoxification, question answering, dialogue. Tricks for training and inference.	2	2	2	2
Uncertainty	Uncertainty estimation for Transformers and NLP	2	2	2	2
Multimodal	Multimodal and vision Transformers	2	2	2	2
Invited lecture	A guest invited lecture on a cutting edge topic in NLP/LLM from a prominent organization in the field	3	2	2	2
Optimization of LLMs	Compression and optimization of LLMs.	2	2	2	2

4. Learning Outcomes

Please specify course intended learning outcomes using [Skoltech Learning Outcomes Framework](#).

1. FUNDAMENTAL KNOWLEDGE

1.2. KNOWLEDGE OF APPLIED SCIENCE AND

2. PERSONAL AND GENERAL PROFESSIONAL SKILLS AND ATTRIBUTES:

2.1. COGNITION AND MODES OF REASONING

2.1.2. System thinking

2.2. ATTITUDES AND LEARNING PROCESS

2.2.1. Initiative and the willingness to take appropriate risks

2.3. ETHICS, EQUITY AND OTHER RESPONSIBILITIES

2.3.1. Ethical action, integrity and courage

3. INTERPERSONAL SKILLS

3.3. COLLABORATION AND CHANGE

3.3.1. Establishing diverse connections and networking

4. LEADING THE INNOVATION PROCESS

4.1. MAKING SENSE OF GLOBAL SOCIETAL ENVIRONMENTAL AND BUSINESS CONTEXT

4.1.1. Appreciating the potential and limitations of science and technology, their role in society and society's role in their evolution

4.2. VISIONING – INVENTING NEW TECHNOLOGIES THROUGH RESEARCH

4.2.1. The research process – hypothesis, evidence and defense

4.2.3. Research aimed at developing new technologies

4.3. VISIONING – CONCEIVING AND DESIGNING SUSTAINABLE SYSTEMS

4.3.2. Identifying and formulating objectives and goals

5. Assignments and Grading

In-person Attendance Requirement 60

In-person Attendance Requirement Comment

Students are expected to attend classes and answer quiz during the lecture.

Assignment Type	Assignment Summary	% of Final Course Grade
Final Exam	Answer in the form of a quiz a set of question related to topics of the course.	20
Test/Quiz	Quizzes during the lectures	20
Homework Assignments	Two individual practical homework assignments based on solution of various NLP shared tasks	60

6. Assessment Criteria

Select Assignment 1 Type

Homework Assignments

Input Sample of Assignment 1 or Share a Link to Assignment 1

Text detoxification

Or Upload Sample of Assignment 1

Assessment Criteria for Assignment 1

- Results (overcoming a baseline, top 1 / top 20% place in the leaderboard)
- Code (readability, reproducibility)
- Report (discussion of results, description of the proposed method)

Select Assignment 2 Type

Homework Assignments

Input Sample of Assignment 2 or Share a Link to Assignment 2

Text categorisation

Or Upload Sample of Assignment 2

Assessment Criteria for Assignment 2

- Results (overcoming a baseline, top 1 / top 20% place in the leaderboard)
- Code (readability, reproducibility)
- Report (discussion of results, description of the proposed method)

Select Assignment 3 Type

Final Exam

Input Sample of Assignment 3 or Share a Link to Assignment 3

1. The scaled dot-product attention is computed as following:
 - $\text{softmax}(QK^T/\sqrt{\text{dim}})V$
 - $\text{sigmoid}(QK^T/\sqrt{\text{dim}})V$
 - $\text{softmax}(QV^T)/K$
 - $\text{softmax}(QV^T/\sqrt{\text{dim}})K$
2. Why masking is needed in the decoder part of Transformer?
 - To hide tokens the model needs to predict
 - To eliminate the computation of duplicate values
 - To allow for additional regularisation though masking (similarly to Dropout)
3. The $O(n^2)$ complexity of the original Attention mechanism where n is:
 - Number of input BPE tokens
 - Number of input words
 - Number of input sentences

Or Upload Sample of Assignment 3

Assessment Criteria for Assignment 3

Correct answer on the quiz questions + fluent answer on the additional question of the lecturer.

Select Assignment 4 Type

Test/Quiz

Input Sample of Assignment 4 or Share a Link to Assignment 4

- 1. Is BERT model an encoder, encoder-decoder, or decoder-based Transformer?
- 2. Is GPT-2 model autoregressive or not?
- 3. Are BPE tokens somehow linguistically motivated?

Assessment Criteria for Assignment 4

True/false answer on the questions given provided set of predefined options.

7. Textbooks and Internet Resources

You can request at most two required textbooks. Additionally, you can suggest up to nine recommended textbooks.

Required Textbooks	ISBN-13 (or ISBN-10)
Dan Jurafsky and James H. Martin (2026): Speech and Language Processing. https://web.stanford.edu/~jurafsky/slp3/	
Foundations of Large Language Models (2025): Tong Xiao, Jingbo Zhu	

8. Facilities

Software
Google Colab
Equipment
A server with GPU is good to have.

Labs for Education

Natural Language Processing Laboratory

9. Additional Notes

Is this syllabus complete?

Yes, the syllabus is a final draft waiting for approval by Education Department

The proposed course 1) has explicit academic content and requirements for receiving credits, 2) is in alignment with the program's learning outcomes, 3) adheres to policies and Skoltech regulations.

Lead Instructor confirms

