МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЛУЦЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
ФАКУЛЬТЕТ РОБОТОТЕХНІКИ ТА ШТУЧНОГО ІНТЕЛЕКТУ

ЗАТВЕРДЖУЮ
Декан факультету
_____ Анатолій ТКАЧУК
«28» серпня 2026 р.

ЗАТВЕРДЖУЮ
Перший проректор
_____ Надія КОВАЛЬЧУК
«29» серпня 2026 р.

СИЛАБУС

з освітнього компонента «Математична та обчислювальна лінгвістика»
ступінь вищої освіти – бакалавр
галузь знань – 11 Математика та статистика
спеціальність – 113 Прикладна математика
освітня програма – Штучний інтелект та аналіз масивів даних

Кафедра штучного інтелекту та математичного моделювання

Силабус склав(ла) доцент кафедри штучного інтелекту та математичного моделювання, к.т.н., доцент Олексій ПРИХОДЬКО

Силабус затверджений на засіданні кафедри штучного інтелекту та математичного моделювання, протокол № 1 від «27» серпня 2026 р.

Завідувач кафедри д.т.н., професор Олена МІКУЛІЧ

Розглянуто та схвалено групою забезпечення освітньої програми «Штучний інтелект та аналіз масивів даних», протокол № 1 від «26» серпня 2026 р.

Гарант ОП «Штучний інтелект та аналіз масивів даних» к.т.н., доцент Олексій ПРИХОДЬКО

Схвалено науково-методичною радою університету, протокол № 1/а від «29» серпня 2026 р.

Затверджений силабус онлайн:
https://e.lntu.edu.ua/syllabus/2306

Луцьк – 2026

СИЛАБУС ОСВІТНЬОГО КОМПОНЕНТА «Математична та обчислювальна лінгвістика»

Освітня програма	Штучний інтелект та аналіз масивів даних
Галузь знань / Спеціальність	11 Математика та статистика / 113 Прикладна математика
Рівень вищої освіти	перший (бакалаврський)
Статус дисципліни	вибіркова
Мова викладання	українська
Обсяг	5 кредитів ECTS / 150 годин
Форма контролю	залік
Викладач(і)	к.т.н., доцент Олексій ПРИХОДЬКО електронна пошта: o.prykhodko@lntu.edu.ua

1. Мета і завдання дисципліни

Мета: Формування у здобувачів теоретичних знань і практичних навичок математичного та алгоритмічного опрацювання природної мови, побудови моделей представлення, аналізу, класифікації та генерації тексту, а також розуміння принципів роботи сучасних нейромережевих і трансформерних мовних моделей

Завдання: сформувати розуміння основних методів попередньої обробки та статистичного аналізу текстових даних;
опанувати ймовірнісні моделі тексту, N-грамні моделі, ланцюги Маркова та методи класифікації текстів;
навчити представляти текст у вигляді векторів із використанням Bag of Words, TF-IDF, LSA та Word Embeddings;
сформувати навички застосування методів лінійної алгебри, теорії ймовірностей і машинного навчання до задач NLP;
ознайомити з принципами роботи нейронних мереж, RNN, механізму уваги та архітектури Transformer;
сформувати розуміння принципів роботи BERT, GPT та великих мовних моделей;
ознайомити з проблемами галюцинацій, підходами RAG, етичними аспектами та сучасними напрямами розвитку NLP.

Пререквізити: Лінійна алгебра та теорія чисел, Теорія ймовірності, математична статистика та ймовірнісні процеси, Python in Datascience, Машинне навчання

2. Компетентності

ФК3 — Здатність обирати та застосовувати математичні методи для розв’язання прикладних задач, моделювання, аналізу, проектування, керування, прогнозування, прийняття рішень.

ФК9 — Здатність до проведення математичного і комп’ютерного моделювання, аналізу та обробки даних, обчислювального експерименту, розв’язання формалізованих задач за допомогою спеціалізованих програмних засобів.

ФК14 — Здатність сформулювати математичну постановку задачі, ґрунтуючись на постановку певної предметної галузі, та обирати метод її розв’язання, що забезпечує потрібні точність і надійність результату.

ФК17 — Здатність до ефективного використання методів та технологій штучного інтелекту при розробці інформаційних інтелектуальних систем, в тому числі аналізу масивів даних

ФК18 — Здатність до вибору адекватних моделей машинного навчання щодо ефективного вирішення конкретних задач аналізу та глибинного добування масивів даних

3. Структура дисципліни, розподіл годин, зв'язок тем з компетентностями

№	Назва теми	Лекції	Практ./ Семін./ лаб	Самост. робота	Разом	Методи навчання	Методи оцінювання	СК
Змістовий модуль 1. Статистика і лінійна алгебра в текстах								
1	Тема 1. Вступ. Як комп'ютер бачить текст	2	2	6	10	МН1, МН2, МН3, МН6	МО2, МО4, МО7, МО9	ФК3, ФК9, ФК14
2	Тема 2. Ймовірність у тексті: N-грамні моделі	2	2	6	10	МН1, МН2, МН3, МН6	МО2, МО4, МО7, МО9	ФК3, ФК9, ФК14
3	Тема 3. Ланцюги Маркова та генерація тексту	2	2	6	10	МН1, МН2, МН3, МН6	МО2, МО4, МО7, МО9	ФК3, ФК9, ФК14
4	Тема 4. Векторний простір тексту (Bag of Words)	2	2	6	10	МН1, МН2, МН3, МН6	МО2, МО4, МО7, МО9	ФК3, ФК9, ФК14
5	Тема 5. Формула TF-IDF та міри близькості	2	2	6	10	МН1, МН2, МН3, МН6	МО2, МО4, МО7, МО9	ФК3, ФК9, ФК14
Змістовий модуль 2. Машинне навчання для текстів (Класика)								
6	Тема 6. Класифікація текстів. Наївний Баєс	2	2	6	10	МН1, МН2, МН3, МН6	МО2, МО4, МО7, МО9	ФК9, ФК14, ФК17, ФК3
7	Тема 7. Матричні розклади та Латентно-семантичний аналіз (LSA)	2	2	6	10	МН1, МН2, МН3, МН6	МО2, МО4, МО7, МО9	ФК3, ФК9, ФК14, ФК17
8	Тема 8. Тематичне моделювання (Topic Modeling)	2	2	6	10	МН1, МН2, МН3, МН6	МО2, МО4, МО7, МО9	ФК3, ФК9, ФК14, ФК17
9	Тема 9. Дистрибутивна семантика (Вступ до Word Embeddings)	2	2	6	10	МН1, МН2, МН3, МН6	МО2, МО4, МО7, МО9	ФК9, ФК14, ФК17, ФК18
10	Тема 10. Алгебра слів (Word2Vec на практиці)	2	2	6	10	МН1, МН2, МН3, МН6	МО2, МО4, МО7, МО9	ФК14, ФК17, ФК18, ФК9
Змістовий модуль 3. Нейронні мережі та Сучасний ШІ								

№	Назва теми	Лекції	Практ./ Семін./ лаб	Самост. робота	Разом	Методи навчання	Методи оцінювання	СК
11	Тема 11. Від класики до нейромереж	2	2	6	10	МН1, МН2, МН3, МН6	МО2, МО4, МО7, МО9	ФК9, ФК14, ФК17, ФК18
12	Тема 12. Механізм Уваги (Attention) та Трансформери (Теорія)	2	2	6	10	МН1, МН2, МН3, МН6	МО2, МО4, МО7, МО9	ФК14, ФК17, ФК18, ФК9
13	Тема 13. Використання Трансформерів (Практика)	2	2	6	10	МН1, МН2, МН3, МН6	МО2, МО4, МО7, МО9	ФК9, ФК14, ФК17, ФК18
14	Тема 14. Великі мовні моделі (LLM) - ChatGPT зсередини	2	2	6	10	МН1, МН2, МН3, МН6	МО2, МО4, МО7, МО9	ФК14, ФК17, ФК18, ФК9
15	Тема 15. Галюцинації ШІ, етика та що буде далі	2	2	6	10	МН1, МН2, МН3, МН6	МО2, МО4, МО7, МО9	ФК17, ФК18
	Разом	**30**	**30**	**90**	**150**			

Легенда: СК — спеціальна (фахова) компетентність (номер за п. 2). Повне формулювання кожного коду наводиться в ОП / стандарті спеціальності.

Методи навчання: МН1 — словесний метод (лекція, дискусія, співбесіда тощо); МН2 — практичний метод (лабораторні та практичні заняття); МН3 — наочний метод (метод ілюстрацій і метод демонстрацій); МН6 — самостійна робота (розв'язання завдань).

Методи оцінювання: МО2 — усне або письмове опитування; МО4 — тестування; МО7 — презентації результатів виконаних завдань та досліджень; МО9 — залік.

4. Результати навчання

Знати:

– основні методи попередньої обробки, статистичного аналізу та представлення текстових даних;

– принципи побудови N-грамних моделей, ланцюгів Маркова та ймовірнісних моделей тексту;

– методи векторного представлення тексту: Bag of Words, TF-IDF, LSA, Word2Vec;

– принципи класифікації текстів із використанням наївного Баєса та нейронних мереж;

– загальні принципи роботи RNN, механізму Attention та архітектури Transformer;

– принципи функціонування моделей BERT, GPT і великих мовних моделей;

– причини виникнення галюцинацій, основи RAG та етичні аспекти використання мовних моделей.

Вміти:

– виконувати токенізацію, стемінг, лематизацію та статистичний аналіз текстових даних;

– будувати та аналізувати N-грамні й марковські моделі тексту;

– формувати векторні представлення документів і слів та оцінювати їхню подібність;

– застосовувати TF-IDF, LSA, Word2Vec та базові методи тематичного моделювання;

– розв'язувати задачі класифікації, аналізу тональності та обробки текстових послідовностей;

– використовувати трансформерні моделі для типових задач NLP;

– аналізувати результати роботи великих мовних моделей, оцінювати їхні обмеження та

доцільність застосування.

5. Цілі сталого розвитку

ЦСР 4 – Якісна освіта

ЦСР 5 – Гендерна рівність

ЦСР 9 – Промисловість, інновації та інфраструктура

6. Неформальна освіта

Рекомендовані курси, тренінги, стажування та інші: результати навчання, здобуті у формальній, неформальній чи інформальній освіті, можуть бути визнані відповідно до процедур ЛНТУ. Підтвердженнями можуть бути сертифікати тематичних тренінгів, вебінарів чи онлайн-курсів разом з описом досягнутих результатів навчання та, за потреби, коротке верифікаційне завдання або портфоліо.

7. Критерії та форми контролю

Оцінювання — за 100-бальною шкалою (національна шкала / ЄКТС). Підсумковий бал = поточний контроль (50 %) + модульний контроль (50 %).

Академічна доброчесність: самостійне виконання завдань; коректне посилання на джерела; позначення частин тексту, згенерованих засобами ШІ. Порушення академічної доброчесності не толеруються.

8. Зміст тем

Тема 1. Вступ. Як комп'ютер бачить текст

Закон Ципфа (Zipf's law).
Проблема розбиття тексту: токенізація, стемінг та лематизація (як очистити текст від зайвого).

Тема 2. Ймовірність у тексті: N-грамні моделі

Ймовірність появи наступного слова.
Що таке біграми та триграми.
Як за допомогою частоти словосполучень комп'ютер вгадує наступне слово на клавіатурі смартфона.

Тема 3. Ланцюги Маркова та генерація тексту

Матриці переходів (Ланцюги Маркова).
Як імовірність залежить лише від поточного стану

Тема 4. Векторний простір тексту (Bag of Words)

Як перетворити документ на математичний вектор.
Модель "Мішок слів" (Bag of Words).
Проблема розріджених матриць (багато нулів)

Тема 5. Формула TF-IDF та міри близькості

Чому "Мішок слів" працює погано.
Формула TF-IDF (Term Frequency - Inverse Document Frequency) - штрафуємо часті слова, заохочуємо рідкісні.
Косинусна міра близькості векторів (чиста лінійна алгебра)

Тема 6. Класифікація текстів. Наївний Баєс

Теорема Баєса. Як вона застосовується для спам-фільтрів.
Проблема нульової ймовірності та згладжування Лапласа (чому не можна множити на нуль)

Тема 7. Матричні розклади та Латентно-семантичний аналіз (LSA)

Сингулярний розклад матриці (SVD) з курсу лінійної алгебри.
Як математика знаходить "приховані" (латентні) зв'язки між словами, стискаючи матрицю TF-IDF

Тема 8. Тематичне моделювання (Topic Modeling)

Логічне продовження LSA.
Як інтерпретувати отримані після стискання вектори як "Теми"

Тема 9. Дистрибутивна семантика (Вступ до Word Embeddings)

Ідея "слово визначається його оточенням".
Перехід від розріджених матриць до щільних векторів. Загальна концепція Word2Vec

Тема 10. Алгебра слів (Word2Vec на практиці)

Щільні векторні представлення слів та їхні семантичні властивості.
Подібність слів у векторному просторі, арифметичні операції над векторами та пошук семантичних аналогій

Тема 11. Від класики до нейромереж

Обмеження класичних методів машинного навчання в задачах обробки тексту.
Багатошаровий перцептрон для класифікації текстів.
Загальні принципи роботи рекурентних нейронних мереж (RNN) та обробки послідовностей

Тема 12. Механізм Уваги (Attention) та Трансформери (Теорія)

Принцип роботи механізму Attention, поняття Query, Key і Value.
Self-Attention та Multi-Head Attention.
Архітектура Transformer, позиційне кодування, паралельна обробка послідовностей та переваги порівняно з RNN

Тема 13. Використання Трансформерів (Практика)

Сімейство моделей BERT.
Принцип двонапрямленого контекстного представлення тексту.
Застосування трансформерів для класифікації, аналізу тональності, розпізнавання сутностей та інших задач NLP

Тема 14. Великі мовні моделі (LLM) - ChatGPT зсередини

Принцип авторегресійної генерації тексту в GPT-моделях.
Передбачення наступного токена, зв’язок із класичними мовними моделями.
Основи промпт-інжинірингу та взаємодії з великими мовними моделями.

Тема 15. Галюцинації ШІ, етика та що буде далі

Причини виникнення галюцинацій у великих мовних моделях. Імовірнісна природа генерації тексту.
Загальні принципи RAG як поєднання пошуку та генерації.
Етичні аспекти використання ШІ та його вплив на майбутні професійні практики

9. Опис ІЗ/КПІЗ

Не передбачено навчальним планом.

10. Загальні вимоги

Освітній процес регулюється академічними правилами ЛНТУ. Здобувачі освіти зобов’язані ознайомитися з вимогами, визначеними Кодексом академічної доброчесності ЛНТУ та Політикою етичного використання штучного інтелекту в ЛНТУ, і дотримуватися принципів академічної доброчесності та етичного використання штучного інтелекту в освітній діяльності.

Академічна доброчесність. Учасники освітнього процесу зобов’язані дотримуватися Закону України «Про академічну доброчесність» від 18.12.2025 № 4742-IX та внутрішніх нормативних документів ЛНТУ. Не допускаються академічний плагіат, фабрикація, фальсифікація, привласнення авторства, несамостійне виконання завдань, а також недоброчесне використання результатів, згенерованих штучним інтелектом.

Використання штучного інтелекту. Інструменти штучного інтелекту можуть використовуватися для пошуку ідей, пояснення методів, попереднього структурування, мовного

редагування та допоміжного аналізу неконфіденційних даних, якщо конкретне завдання не передбачає повністю самостійного виконання. Здобувач освіти має зазначати використаний інструмент, мету та етап його застосування, перевіряти факти й розрахунки та самостійно обґрунтовувати висновки.

11. Література

I. Навчально-методичні матеріали та розробки

Праці автора силабуса:

1. Машинне навчання : метод.і вказівки до виконання самостійної роботи для здобувачів першого (бакалаврського) рівня вищої освіти освітньої програми «Штучний інтелект та аналіз масивів даних» спец. «Прикладна математика» денної форми навч. / уклад. О. С. Приходько. Луцьк : ЛНТУ, 2025. 20 с. URL: https://repository.lntu.edu.ua/handle/123456789/1866

Праці інших авторів:

2. Lee R. S. T. Natural Language Processing: A Textbook with Python Implementation. 2nd ed. Singapore : Springer, 2025. 477 p. URL: https://link.springer.com/book/10.1007/978-981-96-3208-4

3. Haralambous Y. A Course in Natural Language Processing. Cham : Springer, 2024. 534 p. URL: https://link.springer.com/book/10.1007/978-3-031-27226-4

4. Lane H., Dyshel M. Natural Language Processing in Action. 2nd ed. Shelter Island : Manning, 2025. 688 p. URL: https://www.manning.com/books/natural-language-processing-in-action-second-edition

5. Gazit L., Ghaffari M. Mastering NLP from Foundations to LLMs. Birmingham : Packt Publishing, 2024. 340 p. URL: https://www.packtpub.com/en-us/product/mastering-nlp-from-foundations-to-llms-9781804619186

6. Tunstall L., von Werra L., Wolf T. Natural Language Processing with Transformers. Revised ed. Sebastopol : O'Reilly Media, 2022. 408 p. URL: https://www.oreilly.com/library/view/natural-language-processing/9781098136789/

II. Наукові праці

Праці автора силабуса:

7. Yushchuk I., Prykhodko O. Analysis of the limitations of machine learning in predicting sports results. Naukovi Notatky. 2025. № (82). P. 107–111. https://doi.org/10.36910/775.24153966.2025.82.17

III. Інформаційні ресурси

8. Hugging Face. LLM Course. URL: https://huggingface.co/learn/llm-course

9. Stanford University. CS224N: Natural Language Processing with Deep Learning. URL: https://web.stanford.edu/class/cs224n/

10. Scikit-learn. Feature Extraction from Text. URL: https://scikit-learn.org/stable/modules/feature_extraction.html

11. Gensim. Tutorials. URL: https://radimrehurek.com/gensim/auto_examples/tutorials/index.html

12. spaCy. Advanced NLP with spaCy. URL: https://course.spacy.io/en/