



Основи обробки природної мови Силабус освітнього компонента

Реквізити навчальної дисципліни

Рівень вищої освіти	Перший (бакалаврський)
Галузь знань	12 «Інформаційні технології»
Спеціальність	122 «Комп'ютерні науки»
Освітня програма	28343 ОПП Системи і методи штучного інтелекту
Статус дисципліни	вибіркова
Форма навчання	очна(денна)
Рік підготовки, семестр	4 курс, весняний семестр
Обсяг дисципліни	120 годин
Семестровий контроль/ контрольні заходи	залік
Розклад занять	Лекція раз на тиждень, практичні: раз на 2 тижні
Мова викладання	Українська
Інформація про керівника курсу / викладачів	Лектор, Практичні: к.т.н. Шаповал Наталія Віталіївна, shapoval.nataliia@iit.kpi.ua
Розміщення курсу	Google Classroom

Програма навчальної дисципліни

1. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

Метою даної дисципліни є підготовка фахівців з комп'ютерних наук, здатних розв'язувати задачі з області обробки природної мови (NLP). Під час вивчення дисципліни формується вміння використовувати методології NLP для вирішення складних проблем пов'язаних з аналізом та генерацією тексту, аналізувати та обробляти великі дані за допомогою передових методів інтелектуального аналізу даних, специфічних для NLP, покращуючи процес прийняття стратегічних рішень. **Метою навчальної дисципліни є формування та закріплення у студентів наступних компетентностей:** ФК 11 Здатність до інтелектуального аналізу даних на основі методів обчислювального інтелекту включно з великими та погано структурованими даними, їхньої оперативної обробки та візуалізації результатів аналізу в процесі розв'язування прикладних задач; ФК 20 Здатність розробляти системи розпізнавання образів та класифікації в різних предметних областях, обґрунтовано вибирати та використовувати алгоритми розпізнавання образів та проводити навчання систем розпізнавання образів; ФК24 Здатність орієнтуватися в сучасних напрямках розвитку ШІ та нових засобах побудови систем штучного інтелекту та знаходити та розробляти новітні ефективні алгоритми. **Програмні результати навчання, на формування та покращення яких спрямована дисципліна:** ПР12 Застосовувати методи та алгоритми обчислювального інтелекту та інтелектуального аналізу даних в задачах класифікації, прогнозування, кластерного аналізу, пошуку асоціативних правил з використанням програмних

інструментів підтримки багатовимірного аналізу даних на основі технологій DataMining, TextMining, WebMining; ПР19 Розробляти системи розпізнавання образів та класифікації в різних предметних областях, обґрунтовано вибирати та використовувати алгоритми розпізнавання образів та проводити навчання систем розпізнавання образів.

2. Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)

Курс “Основи обробки природної мови” є вибірковим курсом. Цей курс спирається на раніше прочитані спеціальні дисципліни в напрямку аналізу даних та штучного інтелекту і дає систематизоване детальне викладання основ обробки природної мови та аналізу тексту. Ця дисципліна має логічні зв'язки з попередніми дисциплінами навчального плану підготовки, зокрема з курсами “Математичний аналіз”, “Математична статистика”, “Алгоритмізація та програмування”, “Методи та системи штучного інтелекту”.

3. Зміст навчальної дисципліни

Розділ 1. Вступ до NLP та базова обробка тексту

Тема 1.1. Вступ до обробки природної мови

Задачі NLP, приклади застосувань, огляд етапів обробки тексту, ресурси (корпуси, мови).

Тема 1.2. Токенізація, стемінг, лематизація

Препроцесинг тексту, розбиття на токени, нормалізація.

Тема 1.3. Регулярні вирази та основи пошуку шаблонів

Використання RegEx для очищення тексту, вилучення патернів.

Тема 1.4. N-грам моделі та статистичний підхід

Ймовірнісні моделі, перплексія, згладжування.

Тема 1.5. Метрики оцінки NLP-моделей

Accuracy, Precision, Recall, F1, перплексія, BLEU, ROUGE.

Розділ 2. Представлення тексту та семантики

Тема 2.1. Семантичні вектори та подання слів

Bag-of-Words, TF-IDF, проблема розрідженості.

Тема 2.2. Вбудовування слів: Word2Vec, GloVe, FastText

Ідеї контекстного навчання слів, векторні простори.

Тема 2.3. Контекстні вбудовування: ELMo, BERT

Перехід від статичних до контекстних представлень.

Тема 2.4. Подібність текстів та кластеризація

Cosine similarity, кластеризація текстів на основі embeddings.

Розділ 3. Нейромережі в NLP

Тема 3.1. Рекурентні мережі (RNN, LSTM, GRU)

Покрокова обробка тексту, генерація, залежності.

Тема 3.2. Архітектура трансформера та механізм уваги

Self-attention, encoder-decoder, позиційне кодування.

Тема 3.3. BERT: архітектура та навчання

Masked Language Modeling, Next Sentence Prediction, інженерія входів.

Тема 3.4. Тонке налаштування моделей для задач класифікації та аналізу настроїв

Fine-tuning BERT та подібних моделей.

Тема 3.5. Генерація тексту та секвенційні моделі

Sequence-to-sequence, GPT, текстові автозаповнювачі.

Розділ 4. Застосування NLP в реальних задачах

Тема 4.1. Машинний переклад

Нейромережеві підходи до перекладу, Transformer, MarianMT.

Тема 4.2. Відповіді на запитання та пошук інформації

Information retrieval vs reading comprehension, retriever + reader.

Тема 4.3. Чат-боти та діалогові системи

Типи чат-ботів, intent recognition, NLU/NLG.

Тема 4.4. Синтез та розпізнавання мовлення (ASR/TTS)

Короткий огляд Speech-to-Text і Text-to-Speech в NLP-контексті.

Розділ 5. Великі мовні моделі та сучасні виклики

Тема 5.1. Вступ до великих мовних моделей (LLMs)

GPT, T5, PaLM, Claude — особливості, архітектура, масштаби.

Тема 5.2. Підходи до роботи з LLM: prompt engineering, fine-tuning, in-context learning

Як "змусити" LLM вирішувати задачі, використання інструкцій.

Тема 5.3. Етичні аспекти: галюцинації, упередження, відповідальне використання

Bias, fairness, explainability.

4. Навчальні матеріали та ресурси

Основна:

1. [JURAFSKY, Daniel; MARTIN, James H. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.](#)
2. [GOLDBERG, Yoav. Neural network methods for natural language processing. Springer Nature, 2022.](#)

Додаткова

1. Статті які будуть додані в класрумi

Навчальний контент

5. Методика опанування навчальної дисципліни(освітнього компонента)

Лекційні заняття містять детальний огляд тем з Розділів 1-5, а також практичні вправи для розв'язку задач, що розглядаються на лекціях. Передбачається попередній розгляд теми та вправ самостійно перед лекцією для кращого засвоєння матеріалу.

Лабораторні роботи

№ з/п	Перелік лабораторних робіт
1	Лабораторна робота №1. Обробка тексту та побудова N-грам моделі <u>Мета роботи</u> – Ознайомитися з базовими методами препроцесингу тексту: токенизація, очищення, лематизація, побудова N-грам та аналіз частоти слів.
2	Лабораторна робота №2. Векторизація тексту і класифікація на основі TF-IDF <u>Мета роботи</u> – Застосувати підходи Bag-of-Words та TF-IDF для представлення текстів. Навчити класифікатор для задачі аналізу настроїв.
3	Лабораторна робота №3. Використання word embeddings для оцінки подібності текстів <u>Мета роботи</u> – Навчитися використовувати pretrained embeddings для обчислення подібності.
4	Лабораторна робота №4. Класифікація текстів з використанням моделі BERT <u>Мета роботи</u> – Навчитися використовувати трансформерну модель для класифікації тексту з fine-tuning.
	Лабораторна робота №5. Генерація відповідей за допомогою великої мовної моделі <u>Мета роботи</u> – Використати LLM (GPT, T5) для генерації тексту, відповіді на питання або перефразування. Дотренування великих мовних моделей.

● Самостійна робота студента

№ з/п	Вид самостійної роботи	Кількість годин СРС
1	Підготовка до лабораторних робіт	18
2	Підготовка до лекцій	6
3	Підготовка до МКР	6
4	Підготовка до заліку	30

6. Контрольна робота

Метою контрольної роботи є закріплення та перевірка теоретичних знань із освітнього компонента, набуття студентами практичних навичок самостійного вирішення задач та складання та компіляції програм.

Модульна контрольна робота (МКР) розбивається на дві. Одна виконується після вивчення Розділів 1-2 та частину розділу 3. Друга носить дослідницький характер та передбачає розгляд статті з NLP та застосуванню рішення зі статті. Перша контрольна робота проводиться у середовищі Google Classroom. Кожен студент отримує індивідуальне завдання, відповідно до якого необхідно вирішити практичну задачу з використання методів NLP та відповісти на питання відкритого типу за темами які вивчалися.

Політика та контроль

● 7. Політика навчальної дисципліни (освітнього компонента)

Система вимог, які викладач ставить перед студентом:

- правила відвідування занять: заборонено оцінювати присутність або відсутність здобувача на аудиторному занятті, в тому числі нараховувати заохочувальні або штрафні бали. Відповідно до РСО даної дисципліни бали нараховують за відповідні види навчальної активності на лекційних та практичних заняттях.
- правила поведінки на заняттях: студент має можливість отримувати бали за відповідні види навчальної активності на лекційних та практичних заняттях, передбачені РСО дисципліни. Використання засобів зв'язку для пошуку інформації на гугл-диску викладача, в інтернеті, в дистанційному курсі на платформі Сікорський здійснюється за умови вказівки викладача;
- політика дедлайнів та перескладань: якщо студент не проходив або не з'явився на МКР (без поважної причини), його результат оцінюється у 0 балів. Перескладання результатів МКР не передбачено;
- політика щодо академічної доброчесності: Кодекс честі Національного технічного університету України «Київський політехнічний інститут» <https://kpi.ua/files/honorcode.pdf> встановлює загальні моральні принципи, правила етичної поведінки осіб та передбачає політику академічної доброчесності для осіб, що працюють і навчаються в університеті, якими вони мають керуватись у своїй діяльності, в тому числі при вивченні та складанні контрольних заходів з дисципліни «Методи та системи штучного інтелекту»;
- при використанні цифрових засобів зв'язку з викладачем (мобільний зв'язок, електронна пошта, переписка на форумах та у соцмережах тощо) необхідно дотримуватись загальноприйнятих етичних норм, зокрема бути ввічливим та обмежувати спілкування робочим часом викладача.
- Використання засобів генеративного штучного інтелекту регулюється [“Політика використання штучного інтелекту для академічної діяльності в КПІ ім. Ігоря Сікорського”](#). Умови використання генеративного ШІ під час виконання завдання оговорюється в самому завданні.

● 8. Види контролю та рейтингова система оцінювання результатів навчання (PCO)

Поточний контроль: домашні завдання, тестування, МКР, виконання та захист лабораторних робіт.

Календарний контроль: провадиться двічі на семестр як моніторинг поточного стану виконання вимог силабусу.

Семестровий контроль: залік.

Умови допуску до семестрового контролю: виконані та захищені лабораторні роботи, семестровий рейтинг не менше 40 балів.

Таблиця відповідності рейтингових балів оцінкам за університетською шкалою:

Кількість балів	Оцінка
95-100	Відмінно
85-94	Дуже добре
75-84	Добре
65-74	Задовільно
60-64	Достатньо
Менше 60	Незадовільно
Менше 30	Не допущено

Загальна рейтингова оцінка студента після завершення семестру складається з балів, отриманих за:

- тестування по лекційному занятті;
- виконання домашніх завдань;
- виконання та захист лабораторних робіт;
- виконання модульної контрольної роботи (МКР);
- відповіді на екзамені.

Якщо підсумкова оцінка виходить не ціла, то округлення здійснюється в меншу сторону. Наприклад 67,3 \approx 67, або 95,8 \approx 95.

Тестування по лекціям	Домашні завдання	Лабораторні роботи	МКР
14	20	41	25

Тестування по матеріалам лекційних занять

Ваговий бал 1. Максимальна кількість балів за тестування – 2 бал * 7 лекцій = 14 балів.

Тестування проводиться у системі дистанційного навчання *Google Classroom* та доступне протягом 5 робочих днів після завершення поточної лекції. У деяких випадках термін проходження тестування може бути продовжений лектором. Тривалість проходження одного тестування – 10 хвилин. Кількість спроб – одна. У деяких випадках, що пов'язані з технічними проблемами студентів, може надатися повторна спроба на окремі тестування.

Кожне тестування містить 1 відкрите запитання та 1 запитання яке може бути різного формату (вибір правильного варіанту з переліку; вірно/невірно; визначити відповідність; тощо).

Критерії оцінювання

- запитання типу «вибір правильного варіанту з переліку», «вірно/невірно», «чисельна відповідь» оцінюються однозначно: вірна відповідь – 1 бал, невірна відповідь – 0 балів;
- відкриті запитання оцінює ШІ з верифікацією викладача. Повна відповідь 1 бал, неправильна відповідь 0 балів, неповна відповідь 0,5 балів.

Домашні завдання

Ваговий бал 2. Максимальна кількість балів за всі домашні завдання – 5 балів * 4 домашні завдання = 20 балів.

Після деяких лекційних занять студенти отримують домашнє завдання, яке необхідно вирішити та надати на перевірку викладачу впродовж 2х тижнів, однак іноді цей час може бути змінений викладачем у деяких конкретних випадках.

Критерії оцінювання

- домашнє завдання вирішено вірно та здано протягом 2-х тижнів – 5 балів;
- домашнє завдання вирішено вірно, але здано протягом більш ніж 2-х тижнів після лекційного заняття – 3 бали;
- домашнє завдання вирішено із незначними помилками та здано протягом 2-х тижнів після лекційного заняття – 4 бали;
- домашнє завдання вирішено із незначними помилками та здано протягом більш ніж 2-х тижнів після лекційного заняття – 1-2 бали;
- домашнє завдання вирішено із значними помилками – повертається на доопрацювання.

Лабораторні роботи

Ваговий бал. Лабораторні роботи 1-4 мають ваговий бал 8, а лабораторна роботи 5 має ваговий бал 9. Максимальна кількість балів за всі лабораторні роботи складає 8 балів * 4 роботи + 9 балів * 1 роботу= 41 бал.

На лабораторних роботах розв'язується практична задача згідно варіанту з написанням коду. Для допуску до поточної лабораторної роботи необхідно завантажити в Google Classroom код програми з всіма виконаними пунктами згідно завдання лабораторної роботи (або посилання на GitHub). Лабораторні роботи 1-5 виконуються командою (використання GitHub для реалізації командної роботи буде перевагою). Для здачі лабораторної визначається строк (пара на якій вона має бути захищена) задачі. За несвоєчасний захист лабораторної роботи іде зниження балів, в залежності від термінів порушення дедлайнів здачі. Запізнення на 1 пару зниження на 10% від максимальної кількості балів за роботу. Запізнення на 2 пари і більше зниження на 20% від максимальної кількості балів за роботу. Строки задачі можуть бути скореговані за наявності об'єктивних та підтверджених проблем у здобувача.

Критерії оцінювання лабораторної роботи з ваговим балом 8:

- вірно виконані поставлені задачі, продемонстрована працездатність програми, вірні відповіді на запитання до захисту, захищено в визначений строк – 8 балів;
- вірно виконані поставлені задачі, продемонстрована працездатність програми,, відповіді на питання до захисту мають неточності, та/або захищено з запізненням на 1-2 пари – 6-7 балів;
- виконані поставлені задачі, але деякі з них містять помилки або неточності, продемонстрована працездатність не всієї програми, відповіді на питання до захисту мають неточності, та/або захищено з запізненням на 1-2 пари – 3-5 балів;
- виконані не всі поставлені задачі, відповіді на питання до захисту мають значні неточності, та/або захищено з запізненням на 1-2 пари – 0-2 бали;
- код програми не представлений або лабораторна робота не захищена– не зараховується.

Критерії оцінювання лабораторної роботи з ваговим балом 9:

- вірно виконані поставлені задачі, продемонстрована працездатність програми, вірні відповіді на запитання до захисту – 9 балів;
- вірно виконані поставлені задачі, продемонстрована працездатність програми,, відповіді на питання до захисту мають неточності, та/або захищено з запізненням на 1-2 пари – 6- 8 балів;
- виконані поставлені задачі, але деякі з них містять помилки або неточності, продемонстрована працездатність не всієї програми, відповіді на питання до захисту мають неточності, та/або захищено з запізненням на 1-2 пари – 2-5 балів;
- виконані не всі поставлені задачі, відповіді на питання до захисту мають значні неточності, та/або захищено з запізненням на 1-2 пари – 0-1 балів;

- код програми не представлений або лабораторна робота не захищена – не зраховується.

УВАГА! Захист всіх лабораторних робіт є умовою допуску до складання заліку. Студенти, що на момент останньої пари перед заліком не захистили лабораторні роботи, не допускаються до основної здачі та готуються до перескладання.

УВАГА! Для допуску до перескладання заліку треба у визначений викладачем термін здати всі заборгованості по лабораторним роботам.

Модульна контрольна робота

Ваговий бал – 25. Модульна контрольна робота (МКР) розбивається на дві та виконується в *Google Classroom* протягом семестру впродовж 1 тижня. Перша МКР має ваговий бал 15 балів, друга МКР має ваговий бал 10 балів.

Критерії оцінювання модульної контрольної роботи з ваговим балом 15:

На модульній контрольній роботі студент відповідає на відкриті питання та виконує практичне завдання з використанням моделей та методів NLP. Відкриті питання оцінюються в 4 бали, програма в 7 балів.

Відкриті питання оцінюються наступним чином:

- повна відповідь на поставлене питання - 4 бали.
- неповна відповідь на поставлене питання, з явними ознаками використання генеративного ШІ 1-3 бали.
- невірна відповідь - 0 балів.

Програма оцінюється наступним чином:

- програма повністю виконує поставлене завдання, виконувана, зроблені відповідні висновки по отриманим результатам - 7 балів.
- програма не повністю виконує поставлене завдання або немає висновків -4-6 балів.
- програма не виконувана, та/або відсутні висновки, не виконані всі поставлені задачі - 1-3 бали.
- завдання не виконане - 0 балів.

Календарний контроль

Календарний контроль базується на поточній рейтинговій оцінці. Умовою позитивної атестації є значення поточного рейтингу студента не менше 50% від максимально можливого на час атестації.

Додаткові (бонусні) бали

Рейтинговою системою оцінювання передбачені додаткові бали за виконання додаткових завдань. Один студент не може отримати більше ніж 10 бонусних балів у семестрі. Бонусні бали можуть бути отримані за додаткові завдання які оголошуються в *Google Classroom* та на лекціях. Додаткові бали отримують тільки ті студенти, які вірно виконали завдання та завантажили свої відповіді у визначений термін. Додаткові завдання зазвичай змагального характеру, бали отримує той хто здасть завдання першим або отримує найкращий результати за визначеними в завданні метриками, умови оголошуються в самому завданні. Кількість балів за додаткові завдання визначає окремо в самому завданні.

Вправи на лекційних заняттях

Ваговий бал 0,5. Вправи проводяться тільки на лекційних заняттях і доступні тільки у спеціально виділений викладачем час. В інший час незалежно від обставин вправи недоступні. Результати виконання вправи демонструються на лекції. Тривалість проходження однієї вправи від 2 до 5 хвилин, в залежності від її складності. Тривалість вправи попередньо озвучується викладачем. Кількість спроб – одна. Після кожної вправи проводиться коротке обговорення її результатів.

Кожна вправа – це написання коду до завдання/теми лекційного заняття.

Критерії оцінювання

- вірна відповідь – 0,5 балів, невірна відповідь – 0 балів;

Форма семестрового контролю – залік

Максимальна сума балів за роботу у семестрі складає 100. Необхідною умовою допуску до заліку виконані та захищені лабораторні роботи, семестровий рейтинг не менше 35 балів.

Залік містить дві складові: теоретичну та практичну. **Теоретична складова** направлена на перевірку набутих в результаті вивчення освітнього компонента знань студентів. **Практична складова** передбачає перевірку набутих студентами умінь реалізовувати вивчені методи та моделі шІ для вирішення практичних завдань та вирішення задач.

Критерії нарахування балів за контрольні заходи:

~ “відмінно”: 95 -100% - здобувач виявив всебічні, систематичні та глибокі знання навчального матеріалу з дисципліни; продемонстрував уміння вільно виконувати всі завдання, передбачені програмою; засвоїв основну та додаткову літературу; проявив творчі здібності в розумінні, в логічному, чіткому, стислому та ясному трактуванні навчального матеріалу; засвоїв взаємозв'язок основних понять дисципліни, їх значення для подальшої професійної діяльності

~ “дуже добре”: 85-94% - здобувач виявив систематичні знання навчального матеріалу з дисципліни вище середнього рівня; продемонстрував уміння добре виконувати всі завдання, передбачені програмою, допустивши незначні помилки; засвоїв основну та додаткову літературу; засвоїв взаємозв'язок основних понять дисципліни, їх значення для подальшої професійної діяльності

~ “добре”: 75-84% - здобувач виявив загалом добрі знання навчального матеріалу при виконанні передбачених програмою завдань, але припустив ряд помітних помилок; засвоїв основну літературу; показав систематичний характер знань з дисципліни; здатний до їх самостійного використання та поповнення в процесі подальшої навчальної роботи і професійної діяльності

~ “задовільно”: 65-74% - здобувач виявив знання основного навчального матеріалу з дисципліни в обсязі, необхідному для подальшого навчання та майбутньої професійної діяльності; ознайомився з основною літературою; впорався з виконанням завдань, передбачених програмою, але припустив значну кількість помилок або недоліків на запитання при співбесіді, тестуванні та при виконанні завдань тощо, принципів з яких може усунути самостійно

~ “достатньо”: 60-64% - здобувач виявив знання основного навчального матеріалу з дисципліни в мінімальному обсязі, необхідному для подальшого навчання та майбутньої професійної діяльності; ознайомився з основною літературою; в основному виконав завдання, передбачені програмою, але припустив помилки у відповіді на запитання при співбесідах, тестуванні та при виконанні завдань тощо, які він може усунути лише під керівництвом та за допомогою викладача

~ “незадовільно”: 30-59% - здобувач мав значні прогалини в знаннях основного навчального матеріалу; допускав принципові помилки при виконанні передбачених програмою завдань, але спроможний самостійно доопрацювати програмний матеріал і підготуватися для перездачі дисципліни

~ “незадовільно”: 0-29% - здобувач не мав знань зі значної частини навчального матеріалу з дисципліни; допускав принципові помилки при виконанні більшості передбачених програмою завдань або не виконував ці завдання

9. Додаткова інформація з дисципліни (освітнього компонента)

Надається можливість зарахування сертифікатів проходження дистанційних чи онлайн курсів за відповідною тематикою у вигляді додаткових балів.

Робочу програму навчальної дисципліни (силлабус):

Складено доцент, к.т.н., Шаповал Наталія Віталіївна

Ухвалено кафедрою штучного інтелекту (протокол № 14 від 24.06.2025)

Погоджено Методичною комісією ННІПСА (протокол № 7 від 25.06.2025)