

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**  
**ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ ІМЕНІ**  
**СЕМЕНА КУЗНЕЦЯ**

**ЗАТВЕРДЖЕНО**

на засіданні кафедри  
інформатики та комп'ютерної  
техніки  
Протокол № 21 від 29.08.2025 р.

**ПОГОДЖЕНО**

Проректор з навчально-методичної  
роботи



Каріна ПЕМАШКАЛО

**МЕТОДИ ІНТЕЛЕКТУАЛЬНОЇ ОБРОБКИ ПРИРОДНОЇ МОВИ**  
**робоча програма навчальної дисципліни (РПНД)**

Галузь знань	<b>Ф "Інформаційні технології"</b>
Спеціальність	<b>Ф6 "Інформаційні системи і технології"</b>
Освітній рівень	<b>перший (бакалаврський)</b>
Освітня програма	<b>"Штучний інтелект"</b>
Статус дисципліни	<b>обов'язкова</b>
Мова викладання, навчання та оцінювання	<b>українська</b>

Розробник:  
к.т.н., доц.

Наталя БРИНЗА

Завідувач кафедри  
інформатики та комп'ютерної  
техніки

Сергій УДОВЕНКО

Гарант програми

Сергій УДОВЕНКО

**Харків**  
**2025**

## ВСТУП

У сучасних умовах цифрової трансформації досвід провідних ІТ-компаній, фінансових установ, телекомунікаційних операторів та розробників програмного забезпечення свідчить, що одним із найбільш ефективних інструментів підвищення конкурентоспроможності є впровадження систем інтелектуальної обробки природної мови. Саме технології NLP забезпечують автоматизацію аналізу великих масивів текстових даних, що генеруються у вигляді електронної кореспонденції, звернень користувачів, відгуків, технічної документації, повідомлень соціальних мереж та внутрішніх корпоративних комунікацій.

Інтелектуальні інформаційні системи, оснащені модулями обробки природної мови, надають керівникам, аналітикам і розробникам повну, релевантну та оперативну інформацію для прийняття обґрунтованих управлінських і стратегічних рішень. Автоматичне вилучення сутностей, аналіз тональності, тематичне моделювання, класифікація запитів і побудова діалогових систем дозволяють оптимізувати бізнес-процеси, підвищити якість сервісу та зменшити витрати на обробку інформації.

Особливого значення набуває застосування методів машинного навчання та глибинних нейронних мереж для створення адаптивних систем підтримки прийняття рішень, персоналізованих рекомендаційних сервісів та інтелектуальних асистентів. Аналіз текстових даних стає основою для формування ефективних маркетингових стратегій, прогнозування поведінки користувачів і розвитку інноваційних продуктів у сфері штучного інтелекту. Таким чином, дисципліна орієнтована на підготовку фахівців, здатних проектувати та впроваджувати сучасні NLP-рішення в архітектуру інформаційних систем різного призначення.

Навчальна дисципліна "Методи інтелектуальної обробки природної мови" є обов'язковою навчальною дисципліною та вивчається згідно з навчальним планом підготовки здобувачів вищої освіти за спеціальністю Ф6 "Інформаційні системи і технології", освітня програма "Штучний інтелект" першого (бакалаврського) рівня усіх форм навчання усіх форм навчання. Програму навчальної дисципліни розроблено у відповідності до освітньо-професійної програми "Штучний інтелект".

Метою дисципліни є формування у здобувачів вищої освіти теоретичних знань і практичних навичок щодо побудови, аналізу та впровадження моделей автоматичної обробки текстових даних на основі статистичних, лінгвістичних та нейромережових підходів. У межах курсу розглядаються питання морфологічного, синтаксичного та семантичного аналізу тексту, методи векторного представлення слів і документів, алгоритми класифікації, кластеризації, тематичного моделювання, а також сучасні трансформерні архітектури та великі мовні моделі.

Завданням вивчення дисципліни є формування системного розуміння теоретичних засад інтелектуальної обробки природної мови; опанування методів лінгвістичного аналізу; вивчення статистичних і нейромережових

моделей представлення текстових даних; набуття практичних навичок застосування алгоритмів класифікації, кластеризації, тематичного моделювання та аналізу тональності; освоєння принципів побудови та інтеграції трансформерних архітектур і великих мовних моделей у структуру інформаційних систем.

Об'єктом навчальної дисципліни є процеси автоматизованої обробки, аналізу та інтерпретації природномовних текстових даних у межах інформаційних систем і цифрових сервісів.

Предметом навчальної дисципліни є методи, моделі, алгоритми та програмні засоби інтелектуальної обробки природної мови, зокрема підходи машинного навчання, глибинного навчання та лінгвістичного моделювання, що застосовуються для побудови інтелектуальних компонентів систем штучного інтелекту.

Програма навчальної дисципліни передбачає навчання в формі лекцій, лабораторних занять та самостійної роботи здобувачів. Для практичного засвоєння основних тем дисципліни лабораторні заняття, індивідуальна робота та консультації проводяться з застосуванням персональних комп'ютерів, локальних мереж і мережі Інтернет в комп'ютерних класах ХНЕУ ім. С. Кузнеця. Всі види занять забезпечуються необхідною надрукованими та електронними методичним матеріалами.

Результати навчання та компетентності, які формує навчальна дисципліна визначено в табл. 1.

Таблиця 1

**Результати навчання та компетентності, які формує навчальна дисципліна**

Результати навчання	Компетентності, якими повинен оволодіти здобувач вищої освіти
ПР 1	ІК, КЗ 1
ПР 2	ІК, КЗ 6
ПР 4	ІК, КЗ 8
ПР 5	ІК, КЗ 8, КС 1, КС 5
ПР 7	ІК, КЗ 8, КС 8
ПР 10	ІК, КЗ 9, КЗ 10
ПР 11	ІК, КС 8
ПР 12	ІК, КС 15, КС 17

де, ІК. Здатність розв'язувати складні спеціалізовані задачі та практичні проблеми в області інформаційних систем та технологій, або в процесі навчання, що характеризуються комплексністю та невизначеністю умов, які потребують застосування теорій та методів інформаційних технологій.

ПР 1. Знати лінійну та векторну алгебру, диференціальне та інтегральне числення, теорію функцій багатьох змінних, теорію рядів, диференціальні рівняння для функції однієї та багатьох змінних, операційне числення, теорію ймовірностей та математичну статистику в обсязі, необхідному для розробки та використання інформаційних систем, технологій та інфокомунікацій, сервісів та інфраструктури організації.

ПР 2. Застосовувати знання фундаментальних і природничих наук, системного аналізу та технологій моделювання, стандартних алгоритмів та дискретного аналізу при розв'язанні задач проєктування і використання інформаційних систем та технологій.

ПР 4. Проводити системний аналіз об'єктів проєктування та обґрунтовувати вибір структури, алгоритмів та способів передачі інформації в інформаційних системах та технологіях.

ПР 5. Аргументувати вибір програмних та технічних засобів для створення інформаційних систем та технологій на основі аналізу їх властивостей, призначення і технічних характеристик з урахуванням вимог до системи і експлуатаційних умов; мати навички налагодження та тестування програмних і технічних засобів інформаційних систем та технологій.

ПР 10. Розуміти і враховувати соціальні, екологічні, етичні, економічні аспекти, вимоги охорони праці, виробничої санітарії, пожежної безпеки та існуючих державних і закордонних стандартів під час формування технічних завдань та рішень.

ПР 11. Демонструвати вміння розробляти техніко-економічне обґрунтування розроблення інформаційних систем та технологій та вміти оцінювати економічну ефективність їх впровадження.

ПР 12. Розробляти, інтегрувати, аналізувати та підтримувати інформаційні системи штучного інтелекту для вирішення прикладних задач класифікації, кластеризації, прогнозування та генерації нових даних.

КЗ 1. Здатність до абстрактного мислення, аналізу та синтезу.

КЗ 6. Здатність до пошуку, оброблення та узагальнення інформації з різних джерел.

КЗ 8. Здатність оцінювати та забезпечувати якість виконуваних робіт.

КЗ 9. Здатність реалізувати свої права і обов'язки як члена суспільства, усвідомлювати цінності громадянського (вільного демократичного) суспільства та необхідність його сталого розвитку, верховенства права, прав і свобод людини і громадянина в Україні.

КЗ 10. Здатність зберігати та примножувати моральні, культурні, наукові цінності і досягнення суспільства на основі розуміння історії та закономірностей розвитку предметної області, її місця у загальній системі знань про природу і суспільство та у розвитку суспільства, техніки і технологій, використовувати різні види та форми рухової активності для активного відпочинку та ведення здорового способу життя.

КС 1. Здатність аналізувати об'єкт проєктування або функціонування та його предметну область

КС 5. Здатність оцінювати та враховувати економічні, соціальні, технологічні та екологічні фактори на всіх етапах життєвого циклу інфокомунікаційних систем.

КС 8. Здатність управляти якістю продуктів і сервісів інформаційних систем та технологій протягом їх життєвого циклу.

КС 15. Здатність аналізувати сучасні тенденції розвитку інформаційних систем штучного інтелекту.

КС 17. Здатність виконувати навчання нейромережевих моделей штучного інтелекту, оцінювати та контролювати їх якість в процесі використання.

## **ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**

### **Зміст навчальної дисципліни**

**Змістовий модуль 1. Теоретичні основи моделювання мови та прикладні аспекти NLP**

**Тема 1. Моделювання мови і мовлення**

**1.1. Вступ до обробки природної мови.**

Дослідження теоретичних засад та практичних методів комп'ютерного аналізу людської мови. Розгляд рівнів мовної архітектури – від фонетики до прагматики, а також еволюція підходів: від детермінованих лінгвістичних правил до сучасних імовірнісних моделей та нейромережевих технологій.

**1.2. Методи обробки елементів тексту.**

Вивчення технологій препроцесингу неструктурованих даних: токенизація, лематизація та стемінг. Розгляд методів морфологічного аналізу, виділення іменованих сутностей та побудови n-грамних моделей. Формування навичок очищення тексту та його нормалізації для подальшого інтелектуального аналізу.

**1.3. Мовні ресурси і завдання NLP. Діалогова система на основі правил.**

Огляд корпусів текстів, онтологій та лінгвістичних баз даних. Аналіз класичних архітектур чат-ботів, що базуються на регулярних виразах та сценарних деревах. Вивчення методів зіставлення зі зразком (pattern matching) для автоматизації відповідей у закритих доменах.

**1.4. Штучний інтелект в контексті NLP. Мовні моделі і великі мовні моделі.**

Аналіз NLP як ключового напрямку штучного інтелекту, що забезпечує когнітивну взаємодію людини та машини. Розгляд питань семантичного розуміння, машинного перекладу та логічного виводу. Математичне моделювання послідовностей символів та слів. Перехід від статистичних моделей до архітектури Transformer. Вивчення механізмів самоуваги та принципів навчання LLM на гігантських масивах даних для генерації тексту, що не відрізняється від людського.

**Тема 2. Застосування NLP-систем**

**2.1. Класифікація текстів.**

Процес автоматичного розподілу текстових документів за заздалегідь визначеними категоріями. Використовує алгоритми машинного навчання та лінгвістичні правила для аналізу змісту, структури та контексту тексту з метою організації та швидкого пошуку інформації.

## **2.2. Анотування тексту.**

Створення коротких описів або резюме змісту текстових документів для полегшення пошуку, сортування та аналізу інформації. Застосовуються методи автоматичного виділення ключових слів, фраз і тематичних блоків тексту.

## **2.3. Вилучення даних.**

Автоматизований процес виявлення та структурованого збирання релевантної інформації з неструктурованих або частково структурованих джерел. Спрямований на підготовку даних для аналітики, побудови моделей та прийняття обґрунтованих рішень.

## **2.4. Автоматизація перекладу.**

Використання комп'ютерних систем і алгоритмів машинного перекладу для перетворення тексту з однієї мови на іншу. Забезпечує швидке, масштабоване і послідовне перекладення текстів із врахуванням контексту та граматичних правил.

## **Змістовий модуль 2. Сучасні парадигми NLP: від великих моделей до інтелектуальних сервісів**

### **Тема 3. Застосування обробки природної мови**

#### **3.1. Машинний переклад.**

Вивчення еволюції систем перекладу: від правил та статистичних методів до нейромережових архітектур. Розгляд методів оцінки якості перекладу, питання збереження контексту, подолання мовних бар'єрів у вузькоспеціалізованих доменах та використання технологій вирівнювання паралельних корпусів.

#### **3.2. Чат-боти та системи діалогу.**

Дослідження архітектур сучасних діалогових агентів: інтент-орієнтовані системи та генеративні моделі. Розгляд механізмів управління діалогом, розпізнавання намірів користувача та генерація природних відповідей. Аналіз платформ для розробки інтелектуальних асистентів та етичні аспекти автоматизації комунікації.

#### **3.3. Розпізнавання мовлення та перетворення тексту в мовлення.**

Аналіз методів обробки акустичних сигналів та синтезу мовлення. Вивчення спектрального аналізу, прихованих Марковських моделей та сучасних нейронних вокодерів. Розгляд алгоритмів вилучення ознак голосу, моделювання просодики та інтонації для створення природного звукового відтворення текстової інформації.

### **Тема 4. Великі мовні моделі**

#### **4.1. Великі мовні моделі.**

Вивчення архітектур надмасштабних нейронних мереж, заснованих на механізмах трансформерів. Розгляд процесів попереднього на гігантських масивах даних та тонкого налаштування. Аналіз можливості контекстного

навчання, системні обмеження та вектори розвитку сучасних LLM у прикладних задачах.

#### **4.2. Метрики оцінки моделей.**

Дослідження кількісних та якісних методів оцінки ефективності NLP-систем. Вивчення класичних метрик близькості текстів (BLEU, ROUGE, METEOR) та семантичних метрик на основі векторних представлень (BERTScore). Розгляд підходів до оцінки збуреності, перплексії мовних моделей та методи тестування систем на специфічних наборах даних (Benchmarks).

#### **4.3. Проблема галюцинацій.**

Аналіз феномену генерації фактично недостовірної, але лінгвістично коректної інформації. Вивчення причин виникнення галюцинацій: від дефіциту навчальних даних до особливостей ймовірного виводу. Розгляд методів мінімізації помилок через використання зовнішніх джерел знань, верифікацію фактів та обмеження креативності моделей.

### **Тема 5. Завдання NLP-систем**

#### **5.1. Пошук шаблонів у тексті.**

Вивчення методів ідентифікації регулярних структур у текстових масивах. Розгляд синтаксису регулярних виразів для валідації та вилучення даних, а також алгоритми нечіткого пошуку та зіставлення за зразком. Аналіз підходів до побудови гнучких правил пошуку сутностей у слабкоструктурованих інформаційних джерелах.

#### **5.2. Частиномовний та синтаксичний аналіз.**

Дослідження структури речення через POS-тегування та побудову дерев залежностей. Вивчення алгоритми визначення граматичних зв'язків між лексемами, що дозволяє системі розуміти роль кожного слова.

#### **5.3. Семантичний аналіз.**

Вивчення методів вилучення смислового значення з лінгвістичних конструкцій. Розгляд підходів до розв'язання лексичної багатозначності, аналіз семантичних ролей та побудова векторних представлень для оцінки концептуальної близькості текстів. Вивчення методів побудови семантичних просторів. Алгоритми Word2Vec та GloVe. Формування знання про те, як алгоритми інтерпретують контекст та сутність повідомлення.

#### **5.4. Генерування тексту засобами III та LLM.**

Аналіз механізмів створення зв'язного тексту на основі заданих параметрів (prompts). Розглядаються стратегії декодування та їх вплив на різноманітність і логіку контенту. Вивчення методів керування стилем, тональністю та структурою згенерованих відповідей у сучасних архітектурах трансформерів.

Перелік лабораторних занять та завдань за навчальною дисципліною наведено в табл. 2

**Перелік лабораторних занять та завдань**

Назва теми та завдання	Зміст
Тема 1. Завдання 1	Базова обробка тексту та мовне моделювання. Реалізувати класичні методи підготовки тексту та побудувати просту статистичну мовну модель
Тема 2. Завдання 2	Класифікація текстів. Побудувати модель автоматичної класифікації текстів
Тема 3. Завдання 3	Машинний переклад та автоматизація перекладу. Реалізувати просту систему перекладу та порівняти підходи
Тема 4. Завдання 4	Порівняння текстів та семантична подібність. Навчитися вимірювати схожість текстів
Тема 5. Завдання 5	Пошук одного тексту в іншому (інформаційний пошук). Реалізувати систему пошуку фрагментів тексту

Перелік самостійної роботи за навчальною дисципліною наведено в табл. 3.

**Перелік самостійної роботи**

Назва завдання	Зміст
Тема 1 - 5	Вивчення лекційного матеріалу
Тема 1 - 5	Підготовка до лабораторних робіт
Тема 1 - 5	Виконання індивідуальних завдань

Кількість годин лекційних, лабораторних занять та годин самостійної роботи наведено в робочому плані (технологічній карті) з навчальної дисципліни.

**МЕТОДИ НАВЧАННЯ**

У процесі викладання навчальної дисципліни для набуття визначених результатів навчання, активізації освітнього процесу передбачено застосування таких методів навчання, як:

- словесні (лекції, теми 1 - 5), елементи проблемних лекції (теми 1 - 5);
- наочні (демонстрації, теми 1 - 5);
- практичні (лабораторні заняття теми 1 - 5).

В умовах змішаної форми навчання подання лекційного матеріалу та/або проведення лабораторних занять та групових та індивідуальних консультацій

відбувається з використанням платформи Zoom, в умовах звичайної аудиторної форми заняття проводяться очно, в аудиторіях та комп'ютерних залах.

## ФОРМИ ТА МЕТОДИ ОЦІНЮВАННЯ

Університет використовує 100 бальну накопичувальну систему оцінювання результатів навчання здобувачів вищої освіти.

Поточний контроль здійснюється під час проведення лекційних, практичних, лабораторних та семінарських занять і має на меті перевірку рівня підготовленості здобувача вищої освіти до виконання конкретної роботи і оцінюється сумою набраних балів: для дисциплін з формою семестрового контролю залік: максимальна сума – 100 балів; мінімальна сума – 60 балів.

**Підсумковий контроль** включає семестровий контроль та атестацію здобувача вищої освіти.

**Семестровий контроль** проводиться у формі заліку.

**Підсумкова оцінка за навчальною дисципліною** визначається: для дисциплін з формою семестрового контролю залік – сумуванням всіх балів, отриманих під час поточного контролю.

Під час викладання навчальної дисципліни використовуються наступні контрольні заходи:

Поточний контроль: звіти з лабораторних робіт (60 балів), контрольні роботи (20 балів), тести (20 балів).

Семестровий контроль: Залік.

Більш детальну інформацію щодо системи оцінювання наведено в робочому плані (технологічній карті) з навчальної дисципліни.

## РЕКОМЕНДОВАНА ЛІТЕРАТУРА

### Основна

1. Згуровський, М. З. Обробка природної мови [Електронний ресурс] : навч. посіб. для здобувачів ступеня бакалавра, магістра які навчаються за спеціальностями 122 Комп'ютерні науки (F3 Комп'ютерні науки) та 124 Системний аналіз (F4 Системний аналіз та наука про дані) / М. З. Згуровський ; КПІ ім. Ігоря Сікорського. – Електрон. текст. дані (1файл: 3,29 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2025. – 229 с. – Режим доступу : <https://ela.kpi.ua/handle/123456789/72807>.

2. Lee, R. S. T. Natural Language Processing: A Textbook with Python Implementation. Singapore : Springer Singapore, 2024–2025. Сучасний підручник з NLP із практичними прикладами на Python (покриття: моделі мови, токенизація, трансформери, чат-боти). DOI: <https://doi.org/10.1007/978-981-96-3208-4>.

3. Kulkarni, A., Shivananda, A., Kulkarni, A. Natural Language Processing Projects: Build Next-Generation NLP Applications Using AI Techniques. Berkeley, CA : Apress, 2021. DOI: <https://doi.org/10.1007/978-1-4842-7386-9>.

4. Brown P. F., Della Pietra S. A., Della Pietra V. J., Mercer R. L. The mathematics of statistical machine translation: Parameter estimation // *Computational Linguistics*. – 1993. – Vol. 19, No. 2. – P. 263–311. – URL: <https://aclanthology.org/J93-2003>.

5. Papineni K., Roukos S., Ward T., Zhu W.-J. BLEU: a method for automatic evaluation of machine translation // *Proceedings of ACL 2002*. – Philadelphia, 2002. – P. 311–318. – DOI: <https://doi.org/10.3115/1073083.1073135>.

6. Mikolov T., Chen K., Corrado G., Dean J. Efficient estimation of word representations in vector space // *Proceedings of ICLR*. – 2013. – URL: <https://arxiv.org/abs/1301.3781>.

7. Bahdanau D., Cho K., Bengio Y. Neural machine translation by jointly learning to align and translate // *Proceedings of ICLR*. – 2015. – URL: <https://arxiv.org/abs/1409.0473>.

8. Vaswani A., Shazeer N., Parmar N. et al. Attention is all you need // *Advances in Neural Information Processing Systems (NeurIPS)*. – 2017. – Vol. 30. – P. 5998–6008. – URL: <https://arxiv.org/abs/1706.03762>.

#### **Додаткова**

9. *Natural Language Processing (NLP) and Machine Learning (ML)–Theory and Applications*. Edited by Florentina Hristea et al. Basle : MDPI Books, 2022. URL: <https://www.mdpi.com/books/reprint/6302-natural-language-processing-nlp-and-machine-learning-ml-theory-and-applications>.

10. Goldberg, Y. *Neural Network Methods for Natural Language Processing*. Cham : Springer, 2017. DOI: <https://doi.org/10.1007/978-3-031-02165-7>

11. Radford A., Wu J., Child R. et al. Language models are unsupervised multitask learners. – OpenAI, 2019. – URL: [https://cdn.openai.com/better-language-models/language\\_models\\_are\\_unsupervised\\_multitask\\_learners.pdf](https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf).

12. Bommasani R. et al. On the opportunities and risks of foundation models. – 2021. – URL: <https://arxiv.org/abs/2108.07258>.

#### **Інформаційні ресурси**

13. D. Jurafsky, J.H. Martin: *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 3rd ed. <https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf>.

14. Hobson Lane. *Natural Language Processing in Action. Understanding, analyzing, and generating text with Python/* Hobson Lane, Cole Howard, Hannes Hapke. – Manning Publications Co., 2019. – 544 p.