



Розроблення застосунків на основі великих мовних моделей та Generative AI

Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	<i>Другий (магістерський)</i>
Галузь знань	<i>F Інформаційні технології</i>
Спеціальність	<i>F2 Інженерія програмного забезпечення</i>
Освітня програма	<i>Інженерія програмного забезпечення мультимедійних систем та смарттехнологій Інженерія програмного забезпечення іммерсивних систем та платформ цифрових двійників</i>
Статус дисципліни	<i>Вибіркова</i>
Форма навчання	<i>Очна (денна)</i>
Рік підготовки, семестр	<i>1 рік підготовки, 2 семестр</i>
Обсяг дисципліни	<i>Лекції: 32 год., лабораторні заняття: 14 год., самостійна робота: 104 год.</i>
Семестровий контроль/ контрольні заходи	<i>Екзамен, модульна контрольна робота, календарний контроль</i>
Розклад занять	<i>Згідно розкладу на весняний семестр поточного навчального року (schedule.kpi.ua)</i>
Мова викладання	<i>Українська</i>
Інформація про керівника курсу / викладачів	<i>Лектор: аспірант, Федорчук Іван Васильович, fedorchuk.ivan@edu.kpi.ua Лабораторні заняття: аспірант, Федорчук Іван Васильович, fedorchuk.ivan@edu.kpi.ua</i>
Розміщення курсу	<i>Google Classroom. Доступ надається зареєстрованим студентам.</i>

Програма навчальної дисципліни

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

Вивчення дисципліни «Розроблення застосунків на основі великих мовних моделей та Generative AI» дозволяє сформувати у здобувачів освіти теоретичні знання та практичні інженерні навички, необхідні для розроблення програмних продуктів на основі великих мовних моделей та генеративного штучного інтелекту.

Метою вивчення дисципліни «Розроблення застосунків на основі великих мовних моделей та Generative AI» є формування у здобувачів теоретичних знань та практичних інженерних навичок для розроблення програмних продуктів на основі великих мовних моделей та генеративного штучного інтелекту.

Предметом дисципліни «Розроблення застосунків на основі великих мовних моделей та Generative AI» є архітектура та принципи обчислень сучасних великих мовних моделей (LLM), методологія prompt-інженерії та інструменти керування контекстом нейромережових моделей, а також розроблення ШІ-застосунків із використанням хмарних та локальних (open-source) моделей через API інтерфейси.

Вивчення дисципліни «Розроблення застосунків на основі великих мовних моделей та Generative AI» сприяє формуванню у здобувачів освіти **фахових компетентностей**, необхідних для

розв'язання практичних задач професійної діяльності, пов'язаних із розробленням програмних продуктів на основі великих мовних моделей та генеративного штучного інтелекту, включаючи інтеграцію нейромережових моделей у прикладне програмне забезпечення через API інтерфейси, а також розроблення систем контекстного пошуку та автономних ШІ-агентів.

– Здатність обґрунтовано обирати хмарні або локальні нейромережові моделі та інтегрувати їх у прикладне програмне забезпечення за допомогою API інтерфейсів для різних задач.

– Здатність застосовувати методи prompt-інженерії та інструменти керування контекстом великих мовних моделей.

– Здатність розробляти системи контекстного пошуку з використанням векторних баз даних.

– Здатність розробляти та розгортати прикладні генеративні сервіси та автономних ШІ-агентів на основі сучасних фреймворків.

Вивчення дисципліни «Розроблення застосунків на основі великих мовних моделей та Generative AI» сприяє формуванню у студентів наступних **програмних результатів навчання (ПРН)** за освітньою програмою:

– Аналізувати, оцінювати і застосовувати на системному рівні сучасні програмні та апаратні платформи для розв'язання складних задач інженерії програмного забезпечення.

– Аналізувати архітектури великих мовних моделей та застосовувати методи prompt-інженерії і керування контекстом для розв'язання прикладних задач.

– Розробляти ШІ-застосунки на основі великих мовних і генеративних моделей, включаючи їх інтеграцію через програмні інтерфейси та побудову систем контекстного пошуку.

– Проектувати та розробляти автономних ШІ-агентів, обґрунтовувати вибір моделей, інструментів, способів керування контекстом та алгоритмів їх взаємодії.

– Проектувати архітектуру ШІ-застосунків та інтегрувати інтелектуальні компоненти у прикладні програмні системи з урахуванням вимог до модульності, взаємодії сервісів та ресурсних обмежень.

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

Для успішного вивчення дисципліни «Розроблення застосунків на основі великих мовних моделей та Generative AI» здобувач повинен володіти знаннями з навчальних дисциплін першого (бакалаврського) рівня освіти спеціальності F2 Інженерія програмного забезпечення та мати досвід командного розроблення програмного забезпечення.

Отримані при засвоєнні дисципліни «Розроблення застосунків на основі великих мовних моделей та Generative AI» теоретичні знання та практичні уміння забезпечують успішне виконання магістерських дисертацій за спеціальністю F2 Інженерія програмного забезпечення.

3. Зміст навчальної дисципліни

Дисципліна «Розроблення застосунків на основі великих мовних моделей та Generative AI» передбачає вивчення таких тем:

Тема 1. Архітектура великих мовних моделей. Prompt-інженерія

Тема 2. Архітектура LLM-застосунків та управління контекстом

Тема 3. Векторні бази даних та системи контекстного пошуку (RAG)

Тема 4. Автономні ШІ-агенти. Мультиагентні системи

Модульна контрольна робота

Екзамен

4. Навчальні матеріали та ресурси

Базова література:

1. *Обробка природних мов та великі мовні моделі [Електронний ресурс] : навч. посіб. для здобувачів ступеня магістра за освіт. програмами «Інтелектуальні сервіс-орієнтовані обчислювання», «Комп'ютерні науки» спец. F3 Комп'ютерні науки / КПІ ім. Ігоря Сікорського ; уклад.: Р. В. Кислий, І. О. Письменний. – Електрон. текст. дані (1 файл: 1,73 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2026. – 238 с. – Назва з екрана. Матеріали знаходяться у вільному доступі.*
2. *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 (1 файл: 3,29 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2025. – 229 с. – Назва з екрана. Матеріали знаходяться у вільному доступі.*
3. *Згуровський, М. З. Системи і методи штучного інтелекту [Електронний ресурс] / М. З. Згуровський, Ю. П. Зайченко ; НАН України, ННК ІПСА КПІ ім. Ігоря Сікорського. – Електрон. текст. дані (1 файл: 22,64 Мбайт). – Київ : Академперіодика, 2025. – 744 с. – Назва з екрана. Матеріали знаходяться у вільному доступі.*

Додаткова література:

1. *Юрчак І. Розуміння великих мовних моделей: майбутнє штучного інтелекту / Ірина Юрчак, Андрій Хіч, Віра Оксентюк // Комп'ютерні системи проектування. Теорія і практика. — Львів : Видавництво Львівської політехніки, 2024. — Том 6. — № 2. — С. 51–60.*
2. *Liu, Y., Lo, S. K., Lu, Q., Zhu, L., Zhao, D., Xu, X., ... & Whittle, J. (2025). Agent design pattern catalogue: A collection of architectural patterns for foundation model based agents. Journal of Systems and Software, 220, 112278.*
3. *Schulhoff, S., Ilie, M., Balepur, N., Kahadze, K., Liu, A., Si, C., ... & Resnik, P. (2024). The prompt report: a systematic survey of prompt engineering techniques. arXiv preprint arXiv:2406.06608.*
4. *Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., & Chadha, A. (2024). A systematic survey of prompt engineering in large language models: Techniques and applications. arXiv preprint arXiv:2402.07927.*
5. *Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. Advances in neural information processing systems, 33, 9459-9474.*
6. *Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., ... & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. arXiv preprint arXiv:2312.10997.*
7. *Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.*

Ресурси для доступу до великих мовних моделей для практичної роботи:

1. *Free Models Router - API Pricing & Providers | OpenRouter. Режим доступу: <https://openrouter.ai/openrouter/free>. Платформа-маршрутизатор, що пропонує єдиний уніфікований API-інтерфейс для взаємодії з безкоштовними нейромережевими моделями від різних провайдерів.*
2. *Free LLM API Directory (2026): Browse 291+ Models | freellm.net. Режим доступу: <https://freellm.net/models/?free=1>. Цей ресурс являє собою каталог, який надає розробникам інформацію та безкоштовні API-ключі для доступу до сучасних великих мовних моделей для тестування та інтеграції.*
3. *Ollama. Режим доступу: <https://ollama.com/>. Інструмент та фреймворк, який дозволяє здобувачам розгортати, налаштовувати та локально запускати open-source великі мовні*

моделі на власних комп'ютерних ресурсах без необхідності постійного підключення до Інтернету.

Навчальний контент

5. Методика опанування навчальної дисципліни (освітнього компонента)

№ з/п	Тип навчального заняття	Опис навчального заняття
Тема 1. Архітектура великих мовних моделей. Prompt-інженерія		
1	Лекція 1. Архітектура сучасних великих мовних моделей (LLM)	Вступ до Generative AI. Основи архітектури Transformer: механізм уваги, кодери та декодери. Базові принципи обчислень та генерації токенів великими мовними моделями. Огляд відкритих та закритих моделей. Завдання на CPC: п. 6, № 1.
2	Лекція 2. Базові принципи Prompt-інженерії	Концептуальні засади взаємодії з моделями: фреймворк AI Fluency від Anthropic. Структурні архітектури побудови промптів: RTF Framework (Role, Task, Format), CIO Framework (Context, Input, Output), RISEN, CREATE. Інженерні настанови та Prompting Best Practices провідних лабораторій ШІ: рекомендації Anthropic, настанови OpenAI, Google та Microsoft. Робота з ролями, стратегії Zero-shot та Few-shot prompting. Завдання на CPC: п. 6, № 2.
3	Лабораторна робота 1. Застосування технік Prompt-інженерії	Завдання: застосувати структурні фреймворки (RTF, CIO) та методи Zero-shot / Few-shot prompting під час взаємодії зі стандартними веб-інтерфейсами (ChatGPT, Claude, Gemini). Виконати аналіз впливу структурування вхідних інструкцій на якість і повторюваність результату. Завдання на CPC: п. 6, № 3.
4	Лекція 3. Розширені архітектури міркування, декомпозиції та саморефлексії LLM	Огляд просунутих технік промптингу: • Thought Generation: Chain-of-Thought (CoT), Tree of Thoughts (ToT), Graph of Thoughts (GoT). • Problem Decomposition Prompting: Least-to-Most Prompting, Plan-and-Solve Prompting, Program-of-Thoughts Prompting (PoT). • Self-Criticism & Refinement: Self-Evaluation, Self-Refine, Chain-of-Verification (CoVe), System 2 Attention (S2A), Rephrase and Respond (RaR), Re-reading (RE2). Завдання на CPC: п. 6, № 4.
Тема 2. Архітектура LLM-застосунків та управління контекстом		
5	Лекція 4. Програмний доступ до моделей. Еволюція API. Хмарний доступ проти локального розгортання	Еволюція API великих мовних моделей: порівняльний аналіз інтерфейсів OpenAI (Completions, Chat Completions, Responses API), Anthropic (Messages API) та Google Gemini API.

		Хмарні провайдери vs локальне розгортання: метрики TTFT (Time to First Token), TPS (Tokens Per Second), вартість та конфіденційність даних. Інструменти локального запуску моделей на базі Ollama та Hugging Face. Завдання на CPC: п. 6, № 5.
6	Лабораторна робота 2. Програмна взаємодія з LLM через API	Налаштувати програмний доступ до хмарних моделей (локальне розгортання як альтернатива) за допомогою прямих REST-викликів через API без застосування фреймворків. Розробити механізм програмного обміну текстовими повідомленнями з LLM. Реалізувати режим потокової передачі повідомлень для відображення згенерованого тексту в реальному часі. Завдання на CPC: п. 6, № 6.
7	Лекція 5. Архітектури систем на основі LLM. Вступ до фреймворку LangChain	Огляд сучасних архітектур LLM-застосунків (Emerging Architectures for LLM Applications): системні компоненти, потоки даних, контекстні вікна, кешування та рівні абстракції. Вступ до екосистеми LangChain: призначення, базові примітиви (Models, Prompts, Output Parsers). Інтеграція інструментів (Tools/Function Calling) для взаємодії із зовнішнім світом. Обробка мультимодальних повідомлень. Завдання на CPC: п. 6, № 7.
8	Лекція 6. Керування пам'яттю і контекстом. Трасування	Стратегії управління пам'яттю моделі: короткострокова і довгострокова пам'ять. Концепція стандартизації взаємодії з контекстом: огляд Model Context Protocol (MCP). Інструмент LangSmith: моніторинг ланцюгів, профілювання витрат токенів та аналіз затримок. Завдання на CPC: п. 6, № 8.
9	Лабораторна робота 3. Створення базового ШІ-застосунку на LangChain із налаштуванням пам'яті	Завдання: обрати предметну область для агентної системи. Ініціалізувати проєкт на базі LangChain та підключити систему трасування LangSmith. Реалізувати діалоговий ланцюжок з механізмом керування короткостроковою пам'яттю та підключенням зовнішніх інструментів (Tools: розрахункові функції, веб-пошук, інше) для виконання функціональних викликів. Завдання на CPC: п. 6, № 9.
Тема 3. Векторні бази даних та системи контекстного пошуку (RAG)		
10	Лекція 7. Векторне представлення даних та векторні бази даних	Концепція векторних представлень у обробці природної мови. Способи вимірювання семантичної близькості. Архітектура та принципи роботи векторних баз даних. Завдання на CPC: п. 6, № 10.

11	Лабораторна робота 4. Розробка та взаємодія з векторною базою даних	Завдання: згенерувати векторні представлення для власного набору даних. Ініціалізувати локальну векторну базу даних, зберегти у ній вектори та реалізувати функціонал базового семантичного пошуку. Завдання на CPC: п. 6, № 11.
12	Лекція 8. Базова архітектура контекстного пошуку RAG	Концепція відповідей на основі приватних даних за технологією RAG. Життєвий цикл RAG: індексація, пошук контексту, аугментація промпту, генерація відповіді моделі. Завдання на CPC: п. 6, № 12.
13	Лекція 9. Обробка даних для RAG: розбиття тексту та завантаження документів	Стратегії та алгоритми розбиття тексту для збереження семантичного змісту. Методи завантаження та парсингу різних форматів вхідних документів. Завдання на CPC: п. 6, № 13.
14	Лабораторна робота 5. Реалізація базового конвеєра RAG та інтеграція векторної БД	Завдання: розширити існуючий застосунок, реалізувавши повноцінний конвеєр RAG. Налаштувати завантаження власних документів, розбиття їх на чанки, пошук контексту у векторній БД та генерацію остаточної відповіді на основі знайдених даних. Завдання на CPC: п. 6, № 14.
15	Лекція 10. Просунуті техніки RAG	Методи покращення якості пошуку в RAG-системах: гібридний пошук (семантичний + ключові слова) та використання моделей Re-ranking для пересортування знайденого контексту. Графи знань. Завдання на CPC: п. 6, № 15.
Тема 4. Автономні ШІ-агенти. Мультиагентні системи		
16	Лекція 11. Архітектура автономних ШІ-агентів	Принципи побудови автономних ШІ-агентів, патерни проєктування. Концепція Human-in-the-Loop (HITL) для ручного підтвердження або коригування дій агента людиною. Використання Middleware для перехоплення, аудиту та модифікації запитів/відповідей. Завдання на CPC: п. 6, № 16.
17	Лабораторна робота 6. Впровадження механізмів HITL та Middleware	Завдання: додати до розробленого ШІ-застосунку шар Middleware для логування або модифікації запитів користувача до моделі. Впровадити механізм Human-in-the-Loop (HITL), який призупинятиме виконання ланцюжка для отримання підтвердження від користувача перед викликом критичної функції. Завдання на CPC: п. 6, № 17.
18	Лекція 12. Розроблення ШІ-агентів за допомогою LangGraph	Побудова графів виконання для агентів за допомогою LangGraph. Управління станом агента, реалізація циклічних процесів міркування та виконання. Завдання на CPC: п. 6, № 18.

19	Лекція 13. Мультиагентні системи	Взаємодія кількох спеціалізованих агентів для розв'язання складних задач. Стратегії маршрутизації, делегування завдань та координації дій між кількома агентами у спільній системі. Завдання на CPC: п. 6, № 19.
20	Лабораторна робота 7. Розроблення мультиагентної системи на основі LangGraph	Завдання: використовуючи фреймворк LangGraph, спроектувати мультиагентну систему, розподіливши декілька вузькоспеціалізованих агентів за зонами відповідальності (наприклад, агент-аналітик, агент-пошукувець із доступом до RAG-конвеєра з попередніх робіт). Створити головного агента-оркестратора, який керуватиме маршрутизацією запитів, делегуватиме завдання відповідним субагентам та формуватиме консолідовану кінцеву відповідь для користувача.Завдання на CPC: п. 6, № 20.
21	Лекція 14. Розгортання та оцінювання RAG систем	Підходи до оцінювання якості генерації та роботи агентів. Підготовка застосунків до розгортання. Завдання на CPC: п. 6, № 21.
22	Лекція 15. ШІ-застосунки: безпека, вартість та інші виклики	Вектори атак (Prompt Injection, Prompt Hacking) та проблеми витоку приватних даних. Оптимізація роботи з моделями: стратегії зниження вартості API-викликів. Кешування промптів. Завдання на CPC: п. 6, № 22.
Модульна контрольна робота		

6. Самостійна робота студента/аспіранта

Дисципліна «Розроблення застосунків на основі великих мовних моделей та Generative AI» ґрунтується на самостійній підготовці до аудиторних занять на теоретичні та практичні теми.

№ з/п	Назва теми, що виносить на самостійне опрацювання	Кількість годин	Література
1	Підготовка до лекції 1	3	баз. 1, баз. 2, дод. ст. 1, дод. ст. 7
2	Підготовка до лекції 2	3	баз. 1, дод. ст. 3, дод. ст. 4
3	Підготовка до лабораторної роботи 1	4	Всі джерела до лекцій №1-2
4	Підготовка до лекції 3	3	баз. 1, дод. ст. 3, дод. ст. 4
5	Підготовка до лекції 4	3	баз. 1, ресурси для доступу до LLM
6	Підготовка до лабораторної роботи 2	4	Всі джерела до

			лекцій №3-4
7	Підготовка до лекції 5	3	баз. 3, додаткова документація
8	Підготовка до лекції 6	3	баз. 3, додаткова документація
9	Підготовка до лабораторної роботи 3	4	Всі джерела до лекцій №5-6
10	Підготовка до лекції 7	3	баз. 1, баз. 3
11	Підготовка до лабораторної роботи 4	4	Всі джерела до лекції №7
12	Підготовка до лекції 8	3	баз. 1, дод. ст. 5, дод. ст. 6
13	Підготовка до лекції 9	3	баз. 1, дод. ст. 6
14	Підготовка до лабораторної роботи 5	4	Всі джерела до лекцій №8-9
15	Підготовка до лекції 10	3	дод. ст. 5, дод. ст. 6
16	Підготовка до лекції 11	3	баз. 3, дод. ст. 2
17	Підготовка до лабораторної роботи 6	4	Всі джерела до лекції №11
18	Підготовка до лекції 12	3	баз. 3, дод. ст. 2, додаткова документація
19	Підготовка до лекції 13	3	баз. 3, дод. ст. 2
20	Підготовка до лабораторної роботи 7	4	Всі джерела до лекцій №12-13
21	Підготовка до лекції 14	3	баз. 1, баз. 3
22	Підготовка до лекції 15	3	баз. 3
23	Підготовка до модульної контрольної роботи	11	Всі джерела до лекцій №1-15
24	Підготовка до екзамену	20	Весь матеріал семестру

Рекомендовані матеріали для ознайомлення під час самостійної роботи студентів:

- *Prompt engineering* [Електронний ресурс] / OpenAI. – Режим доступу: <https://developers.openai.com/api/docs/guides/prompt-engineering?api-mode=responses>. – Назва з екрана.
- *Prompt engineering overview* [Електронний ресурс] / Claude Platform Docs. – Режим доступу: <https://platform.claude.com/docs/en/build-with-claude/prompt-engineering/overview>. – Назва з екрана.
- *Prompt Engineering Guide* [Електронний ресурс]. – Режим доступу: <https://www.promptingguide.ai/>. – Назва з екрана.
- *LangChain overview* [Електронний ресурс] / Docs by LangChain. – Режим доступу: <https://docs.langchain.com/oss/python/langchain/overview>. – Назва з екрана.

- *LangGraph overview* [Электронный ресурс] / Docs by LangChain. – Режим доступа: <https://docs.langchain.com/oss/python/langgraph/overview>. – Назва з екрана.
- *Introducing the Model Context Protocol* [Электронный ресурс] / Claude by Anthropic. – Режим доступа: <https://www.anthropic.com/news/model-context-protocol>. – Назва з екрана.
- *Retrieval Augmented Generation (RAG) with Deep Agents* [Электронный ресурс] / Docs by LangChain. – Режим доступа: <https://docs.langchain.com/oss/python/deepagents/rag>. – Назва з екрана.
- *Advanced RAG on Hugging Face documentation using LangChain* [Электронный ресурс] / Hugging Face. – Режим доступа: https://huggingface.co/learn/cookbook/advanced_rag. – Назва з екрана.
- *LlamaParse Platform Quickstart* [Электронный ресурс] / Developer Documentation. – Режим доступа: <https://developers.llamaindex.ai/llamaparse/>. – Назва з екрана.
- *Weaviate Database* [Электронный ресурс] / Weaviate Documentation. – Режим доступа: <https://docs.weaviate.io/weaviate>. – Назва з екрана.
- *Tracing quickstart* [Электронный ресурс] / Docs by LangChain. – Режим доступа: <https://docs.langchain.com/langsmith/observability-quickstart>. – Назва з екрана.
- *Prompt engineering best practices for 2026* [Электронный ресурс] / Claude by Anthropic. – Режим доступа: <https://claude.com/blog/best-practices-for-prompt-engineering>. – Назва з екрана.
- *Balaji bal. Prompt Architectures: An Overview of structured prompting strategies* [Электронный ресурс] / Balaji bal // Medium. – Режим доступа: <https://medium.com/@balajibal/prompt-architectures-an-overview-of-structured-prompting-strategies-05b69a494956>. – Назва з екрана.
- *ReACT Agent Model* [Электронный ресурс] / Klu. – Режим доступа: <https://klu.ai/glossary/react-agent-model>. – Назва з екрана.
- *Emerging Architectures for LLM Applications* [Электронный ресурс] / Andreessen Horowitz. – Режим доступа: <https://a16z.com/emerging-architectures-for-llm-applications/>. – Назва з екрана.
- *Nguyen, Bill VU. Run Your Own LLM Locally: A Guide to Self-Hosting* [Электронный ресурс] / Bill VU Nguyen // Medium. – Режим доступа: <https://medium.com/@billvunguyen/run-your-own-llm-locally-a-guide-to-self-hosting-1f3fb8f984bb>. – Назва з екрана.
- *The Best Open Source and Open-Weight LLM Models to Run Locally in 2026* [Электронный ресурс]. – Режим доступа: <https://huggingface.co/blog/daya-shankar/open-source-llm-models-to-run-locally>. – Назва з екрана.
- *An Introduction to Vector Databases for Beginners* [Электронный ресурс] / Xomnia. – Режим доступа: <https://xomnia.com/post/an-introduction-to-vector-databases-for-beginners/>. – Назва з екрана.
- *Word embeddings in NLP: A Complete Guide* [Электронный ресурс]. – Режим доступа: <https://www.turing.com/kb/guide-on-word-embeddings-in-nlp>. – Назва з екрана.
- *Home · facebookresearch/faiss Wiki* [Электронный ресурс] / GitHub. – Режим доступа: <https://github.com/facebookresearch/faiss/wiki>. – Назва з екрана.
- *Essential Chunking Techniques for Building Better LLM Applications* [Электронный ресурс] / MachineLearningMastery.com. – Режим доступа: <https://machinelearningmastery.com/essential-chunking-techniques-for-building-better-llm-applications/>. – Назва з екрана.
- *Better RAG 1: Advanced Basics* [Электронный ресурс]. – Режим доступа: <https://huggingface.co/blog/hrishioa/retrieval-augmented-generation-1-basics>. – Назва з екрана.
- *Better RAG 2: Single-shot is not good enough* [Электронный ресурс]. – Режим доступа: <https://huggingface.co/blog/hrishioa/retrieval-augmented-generation-2-walking>. – Назва з екрана.

- *Better RAG 3: The text is your friend* [Електронний ресурс]. – Режим доступу: <https://huggingface.co/blog/hrishioa/retrieval-augmented-generation-3-structure>. – Назва з екрана.
- *Advanced RAG: Increase RAG Quality with ColBERT Reranker and Llamaindex* [Електронний ресурс]. – Режим доступу: <https://www.pondhouse-data.com/blog/advanced-rag-colbert-reranker>. – Назва з екрана.
- *How to Improve RAG Performance: 5 Key Techniques with Examples* [Електронний ресурс] / DataCamp. – Режим доступу: <https://www.datacamp.com/tutorial/how-to-improve-rag-performance-5-key-techniques-with-examples>. – Назва з екрана.

Безкоштовні онлайн-курси для глибшого ознайомлення:

- *Generative AI for Everyone* [Електронний ресурс] / DeepLearning.AI. – Режим доступу: <https://www.deeplearning.ai/courses/generative-ai-for-everyone>. – Назва з екрана.
- *ChatGPT Prompt Engineering for Developers* [Електронний ресурс] / DeepLearning.AI. – Режим доступу: <https://www.deeplearning.ai/courses/chatgpt-prompt-eng>. – Назва з екрана.
- *Claude 101* [Електронний ресурс] / Claude Academy. – Режим доступу: <https://academy.claude.com/courses/claude-101>. – Назва з екрана.
- *Advanced Prompt Engineering* [Електронний ресурс]. – Режим доступу: <https://learnprompting.org/courses/advanced-prompt-engineering>. – Назва з екрана.
- *Quickstart: LangChain Essentials – Python* [Електронний ресурс]. – Режим доступу: <https://academy.langchain.com/courses/langchain-essentials-python>. – Назва з екрана.
- *Introduction to LangChain: Build AI Agents with Python* [Електронний ресурс]. – Режим доступу: <https://academy.langchain.com/courses/foundation-introduction-to-langchain-python>. – Назва з екрана.
- *Foundation: Introduction to LangGraph – Python* [Електронний ресурс]. – Режим доступу: <https://academy.langchain.com/courses/intro-to-langgraph>. – Назва з екрана.
- *LLMs as Operating Systems: Agent Memory* [Електронний ресурс] / DeepLearning.AI. – Режим доступу: <https://www.deeplearning.ai/courses/llms-as-operating-systems-agent-memory>. – Назва з екрана.
- *Retrieval Augmented Generation (RAG)* [Електронний ресурс] / DeepLearning.AI. – Режим доступу: <https://www.deeplearning.ai/courses/retrieval-augmented-generation>. – Назва з екрана.
- *Building Applications with Vector Databases* [Електронний ресурс] / DeepLearning.AI. – Режим доступу: <https://www.deeplearning.ai/courses/building-applications-vector-databases>. – Назва з екрана.
- *Building Agentic RAG with Llamaindex* [Електронний ресурс] / DeepLearning.AI. – Режим доступу: <https://www.deeplearning.ai/courses/building-agentic-rag-with-llamaindex>. – Назва з екрана.
- *Building and Evaluating Advanced RAG* [Електронний ресурс] / DeepLearning.AI. – Режим доступу: <https://www.deeplearning.ai/courses/building-evaluating-advanced-rag>. – Назва з екрана.

Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

Відвідування лекційних занять є обов'язковим. Проте, відсутність на лекційному занятті не передбачає нарахування штрафних балів, оскільки фінальний рейтинговий бал студента формується виключно на основі оцінювання результатів навчання.

Відвідування лабораторних занять може бути епізодичним та за потреби консультації/захисту лабораторних робіт.

Правила поведінки на заняттях: активність, повага до присутніх та дотримання інших норм етичної поведінки, що визначені розділом 2 Кодексу честі КПІ імені Ігоря Сікорського, з яким студенти ознайомлюються на першому курсі.

Дотримання політики академічної доброчесності, що визначена розділом 3 Кодексу честі КПІ імені Ігоря Сікорського, з яким студенти ознайомлюються на першому курсі. Факти порушень політики академічної доброчесності розглядаються відповідно до «Положення про систему запобігання плагіату, фабрикації, фальсифікації в КПІ ім. Ігоря Сікорського».

Правила захисту лабораторних робіт: роботи повинні бути зроблені згідно варіанту здобувача освіти, що визначається відповідно до постановки завдання кожної лабораторної роботи.

Використання засобів штучного інтелекту під час вивчення дисципліни регламентується відповідно до «Політики використання штучного інтелекту для академічної діяльності в КПІ ім. Ігоря Сікорського». В ході виконання лабораторних робіт допускається використання засобів штучного інтелекту для генерування тих частин, що не становлять предмет вивчення дисципліни: генерування вхідних даних, допомога із вивченням літератури, допомога із вибором предметної галузі для виконання робіт. Виконані лабораторні роботи у тій частині, що становить предмет вивчення дисципліни, мають бути результатом власної оригінальної роботи; допускається використання моделей ШІ для її покращення (рецензування коду і тексту, пошук помилок). Використання штучного інтелекту та його ступінь повинні явно декларуватися у звіті з виконаної лабораторної роботи. Використання засобів штучного інтелекту під час проходження модульної контрольної роботи та семестрового контролю суворо заборонено.

Під час вивчення дисципліни студентам можуть рекомендуватися до ознайомлення англомовні джерела. При цьому рівень знань англійської мови, необхідний для ефективної роботи з цими джерелами, не перевищує рівня, визначеного освітньою програмою (рівень B2).

Засвоєння знань та умінь в ході вивчення дисципліни може бути доступним для більшості осіб з особливими освітніми потребами, окрім здобувачів з серйозними вадами зору, які не дозволяють виконувати завдання за допомогою персональних комп'ютерів, ноутбуків та/або інших технічних засобів навіть за умови їх обладнання спеціалізованим програмним забезпеченням (читачами екрану, засобами голосового вводу).

Визнання результатів неформальної та інформальної освіти відбувається відповідно до «Положення про визнання в КПІ ім. Ігоря Сікорського результатів навчання, набутих у неформальній / інформальній освіті» зі прийнятими змінами на період дії правового режиму воєнного стану. Для валідації результатів такого навчання, здобувач освіти повинен надати викладачу дисципліни документи (сертифікати, свідоцтва, посилання тощо), які визначають тематику, обсяги та перелік результатів навчання, набутих під час неформального навчання, а також результати контролю. На основі наданих документів викладач визначає, які складові дисципліни (чи дисципліна цілком) можуть бути зарахованими.

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

Протягом семестру студенти виконують 7 лабораторних робіт. Максимальна кількість балів за одну лабораторну роботу: 5 балів.

Бали нараховуються за:

- якість виконання лабораторної роботи (звіту): 0–2 бали;*
- опитування (тест) під час захисту лабораторної роботи: 0–2 бали;*
- своєчасне представлення роботи до захисту: 0–1 бал.*

Критерії оцінювання якості виконання лабораторної роботи (звіту):

- 2 бали – робота виконана якісно, в повному обсязі;**

- 1 бал – робота виконана в повному обсязі, але містить незначні помилки;
- 0 балів – робота виконана не в повному обсязі або містить суттєві помилки.

Критерії оцінювання опитування із захисту лабораторної роботи:

- 2 бали – відповідь повна, добре аргументована;
- 1 бал – в цілому відповідь вірна, але має недоліки або незначні помилки;
- 0 балів – немає відповіді, відповідь невірна або містить суттєві помилки.

Максимальна кількість балів за виконання та захист лабораторних робіт:

$$R_{\text{Л}} = 7 \text{ лабораторних робіт} \times 5 \text{ балів} = 35 \text{ балів.}$$

Завдання на модульну контрольну роботу складається з 20 тестових питань – 10 питань з однією правильною відповіддю та 10 питань з декількома правильними відповідями. Кожне питання з однією правильною відповіддю оцінюється в 0.5 балу, кожне питання з декількома правильними відповідями оцінюється в 1 бал.

Критерії оцінювання кожного тестового запитання з однією правильною відповіддю:

- 0.5 балу – відповідь правильна;
- 0 балів – немає відповіді або відповідь невірна.

Критерії оцінювання кожного тестового запитання з декількома правильними відповідями:

- 1 бал – обрано всі правильні відповіді та жодної неправильної;
- 0.5 балу – обрано щонайменше 50% від усіх правильних відповідей;
- 0 балів – немає відповіді або всі відповіді невірні.

Максимальна кількість балів за модульну контрольну роботу:

$$R_{\text{МКР}} = 0.5 \text{ бал} \times 10 \text{ тестових питань з однією правильною відповіддю} + 1 \text{ бал} \times 10 \text{ питань з декількома правильними відповідями} = 15 \text{ балів.}$$

Рейтингова шкала з дисципліни дорівнює:

$$R = R_{\text{с}} + R_{\text{екзамен}} = R_{\text{Л}} + R_{\text{МКР}} + R_{\text{екзамен}} = 35 \text{ балів} + 15 \text{ балів} + 50 \text{ балів} = 100 \text{ балів.}$$

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

- На першій атестації студент отримує «зараховано», якщо його поточний рейтинг становить не менше 50% від максимальної кількості балів, яку може отримати студент до першої атестації.
- На другій атестації студент отримує «зараховано», якщо його поточний рейтинг становить не менше 50% від максимальної кількості балів, яку може отримати студент до другої атестації.

Семестровий контроль: екзамен

Умови допуску до семестрового контролю:

При семестровому рейтингу ($R_{\text{с}}$) не менше 30 балів та зарахуванні усіх лабораторних робіт студент має допуск до екзамену. Після складання екзамену виставляється оцінка відповідно до таблиці.

Необхідною умовою допуску до екзамену є виконання та захист лабораторних робіт.

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
Не виконані умови допуску	Не допущено

Робочу програму навчальної дисципліни (силабус):

Складено аспірант, Федорчук І.В.

Ухвалено кафедрою ПЗКС (протокол № 12 від 11.02.2026 р.)

Погоджено Методичною комісією факультету програмних систем та прикладної математики (протокол № 7 від 23.02.2026 р.)