ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

ЗАТВЕРДЖУЮ
Декан факультету інформаційних технологій
_____ проф. Тетяна ГОВОРУЩЕНКО
«_1_» ____09_____ 2025 р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Людиноцентрований штучний інтелект

Галузь знань – F Інформаційні технології
Спеціальність – F3 Комп'ютерні науки
Рівень вищої освіти – Третій (доктор філософії)
Освітньо-професійна програма – Комп'ютерні науки
Обсяг дисципліни – 6 кредитів ЄКТС
Шифр дисципліни – ОФП.01
Мова навчання – українська
Статус дисципліни – обов'язкова (фахової підготовки)
Факультет - Інформаційних технологій
Кафедра – Комп'ютерних наук

Форма здобуття освіти	Курс	Семестр	Загальний обсяг		Кількість годин						Курсовий проект	Курсова робота	Форма семестрового контролю	
					Аудиторні заняття					Самостійна робота, у т.ч. ІРС				
			Кредити ЄКТС	Години	Всього	Лекції	Лабораторні роботи	Практичні заняття	Семінарські заняття				Залік	Іспит
Д	1	1	6	180	50	16	34			130				+
Разом			**6**	**180**	**50**	**16**	**34**			**130**				**1**

Робоча програма складена на основі освітньо-наукової програми «Комп'ютерні науки» за спеціальністю F3 «Комп'ютерні науки».

Робоча програма складена _____ д-р. техн. н., проф., Едуард МАНЗЮК

Схвалено на засіданні кафедри *комп'ютерних наук*

Протокол від 29. 08. 2025 р., №1 . Зав. кафедри _____ проф. Олександр БАРМАК

Робоча програма розглянута та схвалена вченою радою факультету інформаційних технологій

Голова вченої ради факультету _____ проф. Тетяна ГОВОРУЩЕНКО

Хмельницький 2025

2 ЛИСТ ПОГОДЖЕННЯ

Посада	Назва кафедри	Підпис	Ініціали, прізвище
Завідувач кафедри, д-р. техн. наук, проф.	Комп'ютерні науки		Олександр БАРМАК
Гарант освітньо-професійної програми, д-р. техн. н., проф.	Комп'ютерні науки		Едуард МАНЗЮК

3. Пояснювальна записка

Дисципліна «Людиноцентрований штучний інтелект» є однією із дисциплін фахової підготовки і займає провідне місце у підготовці здобувачів третього (освітньо-наукового) рівня вищої освіти рівня вищої освіти, очної (денної) (далі – денної) форми здобуття вищої освіти, які навчаються за освітньо-професійною програмою «Комп'ютерні науки» в межах спеціальності F3 «Комп'ютерні науки».

Пререквізити – вихідна.

Кореквізити – моделювання та інтелектуальна обробка даних (ОФП.03), педагогічна (викладацька) практика (ОФП.05).

Відповідно до освітньої програми дисципліна сприяє забезпеченню:

***компетентностей*:** здатність продукувати нові ідеї, розв'язувати комплексні проблеми у сфері комп'ютерних наук, застосовувати методологію наукової та педагогічної діяльності, а також проводити власне наукове дослідження, результати якого мають наукову новизну, теоретичне та практичне значення (ІК), здатність до абстрактного мислення, аналізу та синтезу (ЗК01), здатність до пошуку, оброблення та аналізу інформації з різних джерел (ЗК02), здатність виконувати оригінальні дослідження, досягати наукових результатів, які створюють нові знання у комп'ютерних науках та дотичних до них міждисциплінарних напрямах і можуть бути опубліковані у провідних наукових виданнях з комп'ютерних наук та суміжних галузей (ФК01), здатність застосовувати сучасні методології, методи та інструменти експериментальних і теоретичних досліджень у сфері комп'ютерних наук, сучасні цифрові технології, бази даних та інші електронні ресурси у науковій та освітній діяльності (ФК02), здатність виявляти, ставити та вирішувати дослідницькі науково-прикладні задачі та/або проблеми в сфері комп'ютерних наук, оцінювати та забезпечувати якість виконуваних досліджень (ФК03), здатність ініціювати, розробляти і реалізовувати комплексні інноваційні проєкти у галузі комп'ютерних наук та дотичні до неї міждисциплінарних проєктах, демонструвати лідерство під час їх реалізації (ФК04), здатність здійснювати науково-педагогічну діяльність у вищій освіті у сфері комп'ютерних наук (ФК05), здатність розробляти та застосовувати методи і засоби штучного інтелекту з людиноцентрованим підходом для здійснення дослідницької та інноваційної діяльності у галузі комп'ютерних наук, використовуючи інтелектуальні засоби обробки інформації та сучасні інструменти для виконання експериментальних і теоретичних досліджень (УК01).

***програмних результатів навчання*:** мати передові концептуальні та методологічні знання з комп'ютерних наук і на межі предметних галузей, а також дослідницькі навички, достатні для проведення наукових і прикладних досліджень на рівні останніх світових досягнень з відповідного напряму, отримання нових знань та/або здійснення інновацій (ПРН01), формулювати і перевіряти гіпотези; використовувати для обґрунтування висновків належні докази, зокрема, результати теоретичного аналізу, експериментальних досліджень і математичного та/або комп'ютерного моделювання, наявні літературні дані (ПРН03), розробляти та досліджувати концептуальні, математичні і комп'ютерні моделі процесів і систем, ефективно використовувати їх для отримання нових знань та/або створення інноваційних продуктів у комп'ютерних науках та дотичних міждисциплінарних напрямах (ПРН04), визначати актуальні наукові та практичні проблеми у сфері комп'ютерних наук, глибоко розуміти загальні принципи та методи комп'ютерних наук, а також методологію наукових досліджень, застосувати їх у власних дослідженнях у сфері комп'ютерних наук та у викладацькій практиці (ПРН08), використовувати методи і засоби штучного інтелекту з людиноцентрованим підходом для здійснення дослідницької та інноваційної діяльності у галузі комп'ютерних наук, використовуючи інтелектуальні засоби обробки інформації та сучасні інструменти для виконання експериментальних і теоретичних досліджень (ПРН12).

Мета дисципліни. Формування у здобувачів компетентностей, необхідних для розробки та застосування методів і засобів штучного інтелекту з людиноцентрованим підходом, забезпечення етичності та відповідальності ШІ систем, розвитку навичок інтеграції принципів справедливості, прозорості та пояснюваності у процеси дослідницької та інноваційної діяльності у галузі комп'ютерних наук з використанням сучасних інструментів експериментальних і теоретичних досліджень.

Предмет дисципліни. Концепції та принципи відповідального штучного інтелекту, методології MLOps для етичного розгортання ШІ систем, техніки виявлення та мітігації упередженості в алгоритмах машинного навчання, методи забезпечення прозорості та пояснюваності ШІ моделей, принципи людиноцентрованого дизайну та психологічні аспекти взаємодії з ШІ, системи управління, аудиту та сертифікації ШІ систем, операціоналізація етичних принципів у практичних рішеннях ШІ.

Завдання дисципліни. Засвоєння фундаментальних принципів відповідального ШІ та етичних викликів сучасних систем, опанування процесів розробки та MLOps методологій для відповідального розгортання ШІ, формування навичок виявлення та усунення упередженості в системах машинного навчання, розвитку здатностей до створення прозорих та пояснюваних ШІ рішень, засвоєння принципів людиноцентрованого дизайну та мультимодальної взаємодії, опанування методологій аудиту та сертифікації ШІ систем, розвитку навичок операціоналізації етичних принципів у практичних застосуваннях.

Результати навчання. Після вивчення дисципліни студент повинен: застосовувати принципи відповідального ШІ та етичні стандарти у розробці інтелектуальних систем; інтегрувати MLOps процеси з принципами етичного ШІ у хмарних середовищах; виявляти та мітигувати різні типи упередженості на всіх етапах досліджень; розробляти та впроваджувати методи пояснюваності та інтерпретації ШІ моделей; проєктувати людиноцентровані інтерфейси з урахуванням психологічних аспектів взаємодії; проводити аудит та сертифікацію ШІ систем відповідно до регулятивних вимог; реалізовувати гармонізацію великих мовних моделей методами SFT;

розробляти етично обумовлені рекомендаційні системи; застосовувати Human-in-the-Loop підходи для автономних систем; операціоналізувати етичні принципи у практичні бізнес-рішення з використанням ШІ технологій.

4. Структура залікових кредитів дисципліни

Назва розділу (теми)	Кількість годин, відведених на:		
	лекції	лабор. роботи	СРС
Тема 1. Теоретичні основи відповідального штучного інтелекту.	2	4	16
Тема 2. Технології та процеси розробки етичних ШІ систем.	2	4	16
Тема 3. Справедливість та недискримінаційність у системах машинного навчання.	2	4	16
Тема 4. Прозорість та пояснюваність інтелектуальних систем.	2	4	16
Тема 5. Людиноцентрований підхід у розробці ШІ.	2	4	16
Тема 6. Управління та контроль якості ШІ систем.	2	4	16
Тема 7. Практична реалізація принципів відповідального ШІ.	2	4	16
Тема 8. Перспективи розвитку та соціальний вплив відповідального ШІ.	2	6	18
Разом:	16	34	130

5. Програма навчальної дисципліни

5.1 Зміст лекційного курсу

Номер лекції	Перелік тем лекцій, їх анотації	Кількість годин
1	**Основи штучного інтелекту та відповідальний підхід.** Концепції штучного інтелекту та його еволюція. Визначення відповідального ШІ та його ключові принципи. Етичні виклики сучасних ШІ систем. Міжнародні ініціативи та стандарти етичного ШІ. Операціоналізація відповідального підходу в розробці ШІ систем. Роль стейкхолдерів у формуванні етичних принципів ШІ. Літ.: [1] Chapter 1, [2] Chapter 1, 2, [3–5]	2
2	**Процеси розробки та MLOps для відповідального ШІ.** Життєвий цикл машинного навчання та його етапи. MLOps як методологія управління ML проєктами. Інтеграція принципів відповідального ШІ в MLOps процеси. Автоматизація та моніторинг ML систем. Хмарні технології для відповідального розгортання ШІ. Виклики та рішення для масштабування ML систем у виробництві. Літ.: [1] Chapter 7, [2] Chapter 19, [6–8]	2
3	**Упередженість, справедливість та дискримінація в ШІ.** Типи упередженості в системах машинного навчання. Джерела bias: дані, алгоритми, людський фактор. Метрики оцінки справедливості: демографічний паритет, рівність шансів. Методи мітигації упередженості на різних етапах ML pipeline. Справедливість у великих мовних моделях. Правові та регулятивні аспекти недискримінаційного ШІ. Літ.: [1] Chapter 4, [2] Chapter 11, [9–11]	2
4	**Прозорість, пояснюваність та інтерфейси ШІ.** Концепції прозорості та пояснюваності в ШІ. Локальні та глобальні методи пояснення (LIME, SHAP). Інтерпретовність моделей машинного навчання. Дизайн користувацьких інтерфейсів для XAI систем. Виклики пояснюваності в комп'ютерному зорі та NLP. Оцінка якості та корисності пояснень для кінцевих користувачів. Літ.: [1] Chapter 5, [2] Chapter 3, 8, [12–14]	2
5	**Людиноцентрований дизайн та психологічні аспекти ШІ.** Принципи людиноцентрованого підход ШІ систем. Психологічний вплив взаємодії з ШІ на користувачів. Мультимодальна людино-комп'ютерна взаємодія. Методи залучення стейкхолдерів у розробку ШІ. Формування різноманітних команд розробки. Емоційні обчислення та афективний ШІ у контексті етики. Літ.: [1] Chapter3, [2] Chapter 10, 20, 22, [15–17]	2
6	**Управління, аудит та сертифікація ШІ систем.** Структури управління відповідальним ШІ на організаційному рівні. Методології аудиту алгоритмів та ШІ систем. Ризик-менеджмент у ШІ проєктах. Регулятивні вимоги та стандарти сертифікації ШІ. Безперервний моніторинг та валідація ШІ систем. Кібербезпека та захист ШІ систем від атак та зловживань. Літ.: [1] Chapter 8, [2] Chapter 16, [18–20]	2

7	**Операціоналізація відповідального ШІ та етичні принципи.** Переведення етичних принципів у практичні рішення. Кейс-стаді успішної імплементації відповідального ШІ. Розкриття використання ШІ у письмовій діяльності. Роль ШІ етики у зниженні витрат та складності систем. Вплив ШІ на бізнес-процеси організацій. Рекомендаційні системи з врахуванням етичних принципів. Літ.: [1] Chapter 3, 10, 11, [2] Chapter 11, [21–23]	2
8	**Майбутнє відповідального ШІ та соціальний вплив.** Прогнози розвитку відповідального ШІ. Вплив ШІ на економічне зростання та ВВП країн. Трансформація наукових досліджень під впливом великих мовних моделей. Майбутні виклики етики ШІ у контексті AGI. Глобальні ініціативи та міжнародне співробітництво у сфері відповідального ШІ. Підготовка фахівців для епохи відповідального ШІ. Літ.: [1] Chapter 9, [2] Chapter 25, 26, [24–26]	2
	Разом :	16

5.2 Зміст лабораторних занять

№ з/п	Тема лабораторного заняття	Кількість годин
1	Реалізація гармонізації великих мовних моделей методом керованого доналаштування. Літ.: [3–5], [27–30]	4
2	Етапи розробки людиноцентрованої моделі машинного навчання за допомогою сервісу спостереження та налагодження. Літ.: [6–8], [31–33]	4
3	Виявлення та зменшення упередженості в моделях штучного інтелекту. Літ.: [9–11], [34–37]	4
4	Розробка методів пояснюваного штучного інтелекту з використанням нормалізованих ансамблів. Літ.: [12–14], [38–42]	4
5	Машинне навчання з участю людини для автономних систем. Літ.: [15–17], [43–47]	4
6	Парето-оптимальна оцінка справедливості в рекомендаційних системах. Літ.: [18–20], [48–52]	6
7	Мітігація упереджень у великих мовних моделях для етично обумовлених рекомендаційних систем. Літ.: [21–23], [53–56]	6
	Разом:	34

5.3 Зміст самостійної (у т.ч. індивідуальної) роботи здобувача вищої освіти

Самостійна робота студентів усіх форм здобуття освіти полягає у систематичному опрацюванні програмного матеріалу з відповідних джерел інформації, підготовці до виконання і захисту лабораторних робіт. Крім цього до послуг студентів сторінка навчальної дисципліни у Модульному середовищі для навчання, де розміщені Робоча програма дисципліни та необхідні документи з її навчально-методичного забезпечення.

Номер тижня	**Вид самостійної роботи**	**Кількість годин**
1–2	Опрацювання теоретичного матеріалу, підготовка до проведення лабораторної роботи та її захисту, а також виконання індивідуального завдання.	16
3-4	Опрацювання теоретичного матеріалу, підготовка до проведення лабораторної роботи та її захисту, а також виконання індивідуального завдання.	16
5-6	Опрацювання теоретичного матеріалу, підготовка до проведення лабораторної роботи та її захисту, а також виконання індивідуального завдання.	16
7-8	Опрацювання теоретичного матеріалу, підготовка до проведення лабораторної роботи та її захисту, а також виконання індивідуального завдання.	16
9-10	Опрацювання теоретичного матеріалу, підготовка до проведення лабораторної роботи та її захисту, а також виконання індивідуального завдання.	16
11-12	Опрацювання теоретичного матеріалу, підготовка до проведення лабораторної роботи та її захисту, а також виконання індивідуального завдання.	16
13-14	Опрацювання теоретичного матеріалу, підготовка до проведення лабораторної роботи та її захисту, а також виконання індивідуального завдання.	16
15-16	Опрацювання теоретичного матеріалу, підготовка до підсумкового контрольного заходу.	18
	Разом	130

На самостійне опрацювання студентів виносяться визначені у методичних рекомендаціях до лабораторних занять та самостійної роботи індивідуальні завдання (ІЗ) (з відповідних тем. Керівництво самостійною роботою та контроль за виконанням індивідуального завдання здійснюється викладачем згідно з розкладом консультацій у позаурочний час.

Вимоги до виконання індивідуального завдання викладені в Модульному середовищі для навчання на сторінці навчальної дисципліни.

6. Технології та методи навчання

Процес навчання з дисципліни ґрунтується на використанні традиційних та сучасних технологій та методів навчання, зокрема: лекції (з використанням методів візуалізації, проблемного навчання, мотиваційних прийомів, інформаційно-комунікаційних технологій); лабораторні заняття: з використанням методів комп'ютерного моделювання, аналіз проблемних ситуацій, пояснення, дискусія тощо, самостійна робота: робота над засвоєнням теоретичного матеріалу, виконання індивідуальних завдань, підготовка до поточного та підсумкового контролю тощо), з використанням інформаційно-комп'ютерних технологій та технологій дистанційного навчання.

7. Методи контролю

Поточний контроль здійснюється під час аудиторних лабораторних занять, а також у дні проведення контрольних заходів, встановлених робочою програмою і графіком освітнього процесу, в т.ч. з використанням Модульного середовища для навчання. При цьому використовуються такі методи поточного контролю:

- оцінювання результатів захисту лабораторних робіт;
- оцінювання індивідуальних завдань.

При виведенні підсумкової семестрової оцінки враховуються результати як поточного контролю, так і підсумкового контролю, який проводиться з усього матеріалу дисципліни за білетами, попередньо розробленими і затвердженими на засіданні кафедри. Здобувач вищої освіти, який набрав з будь-якого виду навчальної роботи, суму балів нижчу за 60 відсотків від максимального балу, ***не допускається*** до семестрового контролю, поки не виконає обсяг роботи, передбачений Робочою програмою. Здобувач вищої освіти, який набрав позитивний середньозважений бал (60 відсотків і більше від максимального балу) з усіх видів поточного контролю і не склав іспит, вважається таким, який ***має*** академічну заборгованість. Ліквідація академічної заборгованості із семестрового контролю здійснюється у період екзаменаційної сесії або за графіком, встановленим деканатом відповідно до «Положення про контроль і оцінювання результатів навчання здобувачів вищої освіти у ХНУ».

8. Політика дисципліни

Політика навчальної дисципліни загалом визначається системою вимог до здобувача вищої освіти, що передбачені чинними положеннями Університету про організацію і навчально-методичне забезпечення освітнього процесу. Зокрема, проходження інструктажу з техніки безпеки; відвідування занять з дисципліни є обов'язковим. За об'єктивних причин (підтверджених документально) теоретичне навчання за погодженням із лектором може відбуватись в он-лайн режимі. Успішне опанування дисципліни і формування фахових компетентностей і програмних результатів навчання передбачає необхідність підготовки до лабораторного заняття (вивчення теоретичного матеріалу з теми роботи, попередню підготовку протоколу роботи, (наведені у Методичних рекомендаціях до лабораторних занять)), активно працювати на занятті, якісно підготувати звіт (протокол роботи відповідно до теми), захистити результати виконаної роботи, брати участь у дискусіях щодо прийнятих конструктивних рішень при виконанні здобувачами лабораторних робіт тощо.

Здобувачі вищої освіти мають дотримуватися встановлених термінів виконання всіх видів навчальної роботи відповідно до робочої програми навчальної дисципліни. Термін захисту лабораторної роботи вважається своєчасним, якщо студент захистив її на наступному після виконання роботи занятті. Пропущене лабораторне заняття студент зобов'язаний відпрацювати у встановлений викладачем термін, але не пізніше, ніж за два тижні до кінця теоретичних занять у семестрі.

Засвоєння студентом теоретичного матеріалу з дисципліни оцінюється за результатами опитування під час захисту лабораторних робіт та контрольної роботи.

Здобувач вищої освіти, виконуючи самостійну роботу або індивідуальну роботу з дисципліни, має дотримуватися політики доброчесності (заборонені списування, плагіат (в т.ч. із використанням мобільних девайсів)). У разі виявлення порушення політики академічної доброчесності в будь-яких видах навчальної роботи здобувач вищої освіти отримує незадовільну оцінку і має повторно виконати завдання з відповідної теми (виду роботи), що передбачені робочою програмою. Будь-які форми порушення академічної доброчесності ***не допускаються.***

У межах вивчення навчальної дисципліни здобувачам вищої освіти передбачено визнання і зарахування результатів навчання, набутих шляхом неформальної освіти, що розміщені на доступних платформах, які сприяють формування компетентностей і поглибленню результатів навчання, визначених робочою програмою дисципліни, або забезпечують вивчення відповідної теми та/або виду робіт з програми навчальної дисципліни (детальніше у Положенні про порядок визнання та зарахування результатів навчання здобувачів вищої освіти у ХНУ).

Окремі теми курсу можуть бути зараховані у випадку отримання студентом **результатів навчання у неформальній освіті**, що підтверджені відповідним документом (сертифікат, свідоцтво, освітня програма тощо) та відповідно до Положення про порядок визнання та перезарахування результатів навчання здобувачів вищої освіти у ХНУ (https://khmnu.edu.ua/wp-content/uploads/normatyvni-dokumenty/polozhennya/pro-poryadok-vyznannya-ta-perezarahuvannya-rezultativ-navchannya.pdf). Перелік лабораторних робіт, що можуть бути перезараховані та

приклади відповідних курсів на освітніх ресурсах:

Лабораторні роботи 1-7: Trustworthy AI: Managing Bias, Ethics, and Accountability– https://www.coursera.org/learn/responsible-ai-and-ethics

Лабораторні роботи 1-7: AI Ethics & Responsible AI: Build Ethical and Trustworthy AI – https://www.udemy.com/course/ai-ethics-responsible-development-and-use-course/?couponCode=MT250915G1

9. Оцінювання результатів навчання студентів у семестрі

Оцінювання академічних досягнень здобувача вищої освіти здійснюється відповідно до «Положення про контроль і оцінювання результатів навчання здобувачів вищої освіти у ХНУ». При поточному оцінюванні виконаної здобувачем роботи з кожної структурної одиниці і отриманих ним результатів викладач виставляє йому певну кількість балів із встановлених Робочою програмою для цього виду роботи. При цьому кожна структурна одиниця навчальної роботи може бути зарахована, якщо здобувач набрав не менше 60 відсотків (мінімальний рівень для позитивної оцінки) від максимально можливої суми балів, призначеної структурній одиниці.

При оцінюванні результатів навчання здобувачів вищої освіти з будь-якого виду навчальної роботи (структурної одиниці) рекомендується використовувати наведені нижче узагальнені критерії:

Таблиця – Критерії оцінювання навчальних досягнень здобувача вищої освіти

Оцінка та рівень досягнення здобувачем запланованих ПРН та сформованих компетентностей	**Узагальнений зміст критерія оцінювання**
Відмінно (високий)	Здобувач вищої освіти глибоко і у повному обсязі опанував зміст навчального матеріалу, легко в ньому орієнтується і вміло використовує понятійний апарат; уміє пов'язувати теорію з практикою, вирішувати практичні завдання, впевнено висловлювати і обґрунтовувати свої судження. Відмінна оцінка передбачає логічний виклад відповіді мовою викладання (в усній або у письмовій формі), демонструє якісне оформлення роботи і володіння спеціальними приладами та інструментами, прикладними програмами. Здобувач не вагається при видозміні запитання, вміє робити детальні та узагальнюючі висновки, демонструє практичні навички з вирішення фахових завдань. При відповіді допустив дві–три несуттєві ***похибки***.
Добре (середній)	Здобувач вищої освіти виявив повне засвоєння навчального матеріалу, володіє понятійним апаратом, орієнтується у вивченому матеріалі; свідомо використовує теоретичні знання для вирішення практичних задач; виклад відповіді грамотний, але у змісті і формі відповіді можуть мати місце окремі неточності, нечіткі формулювання правил, закономірностей тощо. Відповідь здобувача вищої освіти будується на основі самостійного мислення. Здобувач вищої освіти у відповіді допустив дві–три ***несуттєві помилки***.
Задовільно (достатній)	Здобувач вищої освіти виявив знання основного програмного матеріалу в обсязі, необхідному для подальшого навчання та практичної діяльності за професією, справляється з виконанням практичних завдань, передбачених програмою. Як правило, відповідь здобувача вищої освіти будується на рівні репродуктивного мислення, здобувач вищої освіти має слабкі знання структури навчальної дисципліни, допускає неточності і ***суттєві помилки*** у відповіді, вагається при відповіді на видозмінене запитання. Разом з тим, набув навичок, необхідних для виконання нескладних практичних завдань, які відповідають мінімальним критеріям оцінювання і володіє знаннями, що дозволяють йому під керівництвом викладача усунути неточності у відповіді.
Незадовільно (недостатній)	Здобувач вищої освіти виявив розрізнені, безсистемні знання, не вміє виділяти головне і другорядне, допускається помилок у визначенні понять, перекручує їх зміст, хаотично і невпевнено викладає матеріал, не може використовувати знання при вирішенні практичних завдань. Як правило, оцінка «незадовільно» виставляється здобувачеві вищої освіти, який не може продовжити навчання без додаткової роботи з вивчення навчальної дисципліни.

Структурування дисципліни за видами навчальної роботи і оцінювання результатів навчання студентів денної форми здобуття освіти у семестрі

Лабораторна робота							**Самостійна робота**	**Семестровий контроль**	**Разом**
							ІЗ	**Іспит**	Сума балів
1	2	3	4	5	6	7	1		
3-5	3-5	3-5	3-5	3-5	3-5	3-5	15-25	24-40	**60-100***
21-35							**15-25**	**24-40**	

Примітки: *За набрану з будь-якого виду навчальної роботи з дисципліни кількість балів, нижче встановленого мінімуму, здобувач отримує незадовільну оцінку і має її перездати у встановлений викладачем

(деканом) термін. Інституційна оцінка встановлюється відповідно до таблиці «**Співвідношення інституційної шкали оцінювання і шкали оцінювання ЄКТС»**.

Оцінювання результатів захисту лабораторної роботи

Виконана й оформлена відповідно до встановлених методичними рекомендаціями вимог лабораторна робота комплексно оцінюється викладачем при її захисті з урахуванням таких критеріїв: самостійність та правильність виконання; знання теоретичного матеріалу; здатність аргументувати прийняті рішення; повнота та якість отриманих результатів; дотримання вимог до оформлення та структурування роботи; уміння виправляти виявлені помилки у ході захисту.

Результат виконання і захисту здобувачем вищої освіти кожної лабораторної роботи оцінюється відповідно до таблиці "Критеріїв оцінювання навчальних досягнень здобувача вищої освіти" (мінімальний позитивний бал – 3 бали, максимальний – 5 балів).

У випадку виявлення здобувачем рівня знань, нижчого ніж 60 відсотків від максимального балу, встановленого Робочою програмою для кожної структурної одиниці, лабораторна робота йому *не зараховується* і для її захисту він має детальніше опрацювати матеріал з теми роботи, методику її виконання, виправити грубі помилки та повторно вийти на її захист у призначений для цього викладачем час.

Оцінювання результатів виконання індивідуального завдання

Виконане та оформлене відповідно до вимог, визначених Методичними рекомендаціями, індивідуальне завдання (ІЗ) комплексно оцінюється викладачем з урахуванням таких критеріїв: самостійність виконання; правильність розв’язання поставлених задач; обґрунтованість вибору методів розв’язання; повнота пояснень та аргументованість відповідей; якість оформлення та дотримання вимог до структури і змісту роботи.

Результат виконання здобувачем вищої освіти кожного ІЗ оцінюється відповідно до таблиці Критеріїв оцінювання навчальних досягнень здобувача вищої освіти з урахуванням рівня досягнення запланованих програмних результатів навчання та сформованих компетентностей. За підсумками захисту присвоюється відповідна сума балів (мінімальний позитивний бал – 15 балів, максимальний – 25 балів).

У разі, якщо здобувач вищої освіти виявив рівень знань і виконання ІЗ, що нижчий ніж 60 відсотків від максимальної кількості балів, встановленої Робочою програмою для цієї структурної одиниці, завдання не зараховується. У такому випадку студент має повторно опрацювати зміст завдання, усунути помилки та здати на перевірку доопрацьоване ІЗ у терміни, погоджені з викладачем.

Оцінювання результатів підсумкового семестрового контролю (іспит).

Освітня програма передбачає підсумковий семестровий контроль з дисципліни у формі іспиту, завданням якого є системне й об’єктивне оцінювання як теоретичної, так і практичної підготовки здобувача з навчальної дисципліни. Складання іспиту відбувається за попередньо розробленими і затвердженими на засіданні кафедри білетами. Відповідно до цього в екзаменаційному білеті пропонується поєднання питань як теоретичного (в т.ч. у тестовій формі), так і практичного характеру.

Таблиця – Оцінювання результатів підсумкового семестрового контролю здобувачів денної форми навчання (40 балів для підсумкового контролю)

Види завдань	**Для кожного окремого виду завдань**		
	Мінімальний (достатній) бал (задовільно)	**Потенційні позитивні бали* (середній бал) (добре)**	**Максимальний (високий) бал (відмінно)**
Теоретичне питання № 1	3	4	5
Теоретичне питання № 2	3	4	5
Практичне завдання	18	24	30
Разом:	**24**		**40**

Примітка. *Позитивний бал за іспит, відмінний від мінімального (24 бали) та максимального (40 балів), знаходиться в межах 25-39 балів та розраховується як сума балів за усі структурні елементи (завдання) іспиту.

Для кожного окремого виду завдань підсумкового семестрового контролю застосовуються критерії оцінювання навчальних досягнень здобувача вищої освіти, наведені вище (**Таблиця – Критерії оцінювання навчальних досягнень здобувача вищої освіти)**.

Підсумкова семестрова оцінка за інституційною шкалою і шкалою ЄКТС визначається в автоматизованому режимі після внесення викладачем результатів оцінювання у балах з усіх видів навчальної роботи до електронного журналу. Співвідношення інституційної шкали оцінювання і шкали оцінювання ЄКТС наведені нижче у таблиці «Співвідношення».

Семестровий іспит виставляється, якщо загальна сума балів, яку набрав студент з дисципліни за результатами поточного контролю, знаходиться у межах від 60 до 100 балів. При цьому за інституційною шкалою ставиться оцінка «відмінно/добре/задовільно», а за шкалою ЄКТС – буквене позначення оцінки, що відповідає набраній студентом сумі балів відповідно до таблиці Співвідношення.

Таблиця – Співвідношення інституційної шкали оцінювання і шкали оцінювання ЄКТС

<table>
<tr><th rowspan="2">Оцінка ЄКТС</th><th rowspan="2">Рейтингова шкала балів</th><th colspan="2">Інституційна оцінка (рівень досягнення здобувачем вищої освіти запланованих результатів навчання з навчальної дисципліни)</th></tr>
<tr><th>Залік</th><th>Іспит/диференційований залік</th></tr>
<tr><td>A</td><td>90-100</td><td rowspan="5">Зараховано</td><td>Відмінно/Excellent – високий рівень досягнення запланованих результатів навчання з навчальної дисципліни, що свідчить про безумовну готовність здобувача до подальшого навчання та/або професійної діяльності за фахом</td></tr>
<tr><td>B</td><td>83-89</td><td rowspan="4">Добре/Good – середній (максимально достатній) рівень досягнення запланованих результатів навчання з навчальної дисципліни та готовності до подальшого навчання та/або професійної діяльності за фахом
Задовільно/Satisfactory – Наявні мінімально достатні для подальшого навчання та/або професійної діяльності за фахом результати навчання з навчальної дисципліни</td></tr>
<tr><td>C</td><td>73-82</td></tr>
<tr><td>D</td><td>66-72</td></tr>
<tr><td>E</td><td>60-65</td></tr>
<tr><td>FX</td><td>40-59</td><td rowspan="2">Незараховано</td><td>Незадовільно/Fail – Низка запланованих результатів навчання з навчальної дисципліни відсутня. Рівень набутих результатів навчання є недостатнім для подальшого навчання та/або професійної діяльності за фахом</td></tr>
<tr><td>F</td><td>0-39</td><td>Незадовільно/Fail – Результати навчання відсутні</td></tr>
</table>

10. Питання для самоконтролю результатів навчання

1. Яке визначення відповідального штучного інтелекту та які його ключові принципи?
2. Поясніть різницю між AI Alignment та AI Safety. Наведіть приклади кожного підходу.
3. Які основні етичні виклики виникають при розробці сучасних ШІ систем?
4. Охарактеризуйте роль стейкхолдерів у формуванні етичних принципів ШІ. Хто повинен брати участь у цьому процесі?
5. Як операціоналізувати відповідальний підхід на різних етапах розробки ШІ систем?
6. Проаналізуйте ключові положення AI Act Європейського Союзу та їх вплив на розробку ШІ.
7. Які міжнародні ініціативи та стандарти етичного ШІ існують на сьогодні?
8. Опишіть життєвий цикл машинного навчання та його основні етапи з точки зору відповідального ШІ.
9. Що таке MLOps та як інтегрувати принципи відповідального ШІ в MLOps процеси?
10. Які виклики та рішення існують для масштабування ML систем у виробництві з дотриманням етичних принципів?
11. Як здійснювати автоматизацію та моніторинг ML систем з урахуванням етичних аспектів?
12. Опишіть роль хмарних технологій у відповідальному розгортанні ШІ систем.
13. Які метрики та KPI використовуються для оцінки етичності ШІ систем в продакшені?
14. Класифікуйте основні типи упередженості в системах машинного навчання та наведіть приклади кожного.
15. Які джерела bias існують на рівні даних, алгоритмів та людського фактору?
16. Поясніть різницю між демографічним паритетом та рівністю шансів як метриками справедливості.
17. Опишіть методи мітигації упередженості на різних етапах ML pipeline.
18. Як виявляти та усувати справедливість у великих мовних моделях?
19. Які правові та регулятивні аспекти недискримінаційного ШІ слід враховувати?
20. Як забезпечити міжгрупову справедливість при розробці рекомендаційних систем?
21. В чому полягає різниця між прозорістю та пояснюваністю в ШІ системах?
22. Порівняйте локальні та глобальні методи пояснення (LIME, SHAP, тощо).
23. Як забезпечити інтерпретовність моделей машинного навчання для різних типів алгоритмів?
24. Опишіть принципи дизайну користувацьких інтерфейсів для XAI систем.
25. Які специфічні виклики пояснюваності існують у комп'ютерному зорі та NLP?
26. Як оцінити якість та корисність пояснень для кінцевих користувачів?
27. Наведіть приклади успішного впровадження XAI у критично важливих доменах.
28. Сформулюйте основні принципи людиноцентрованого дизайну ШІ систем.
29. Як психологічний вплив взаємодії з ШІ впливає на користувачів та як його врахувати?
30. Опишіть методи мультимодальної людино-комп'ютерної взаємодії в контексті етичного ШІ.
31. Які методи залучення стейкхолдерів у розробку ШІ найбільш ефективні?
32. Чому важливо формувати різноманітні команди розробки ШІ та як це робити?
33. Як емоційні обчислення та афективний ШІ співвідносяться з етичними принципами?
34. Наведіть приклади Human-in-the-Loop підходів для автономних систем.
35. Опишіть структури управління відповідальним ШІ на організаційному рівні.
36. Які методології аудиту алгоритмів та ШІ систем існують та як їх застосовувати?
37. Як здійснювати ризик-менеджмент у ШІ проєктах з урахуванням етичних аспектів?
38. Охарактеризуйте регулятивні вимоги та стандарти сертифікації ШІ систем.
39. Як організувати безперервний моніторинг та валідацію ШІ систем?

40. Які аспекти кібербезпеки та захисту ШІ систем від атак критично важливі?
41. Опишіть процес проведення етичного аудиту ШІ системи.
42. Як перевести етичні принципи у практичні технічні рішення при розробці ШІ?
43. Проаналізуйте кейс-стаді успішної імплементації відповідального ШІ в організації.
44. Як правильно розкривати використання ШІ у письмовій та творчій діяльності?
45. Яка роль ШІ етики у зниженні витрат та складності систем?
46. Опишіть методи реалізації гармонізації великих мовних моделей методом SFT.
47. Як розробляти етично обумовлені рекомендаційні системи?
48. Які прогнози розвитку відповідального ШІ на найближчі 5-10 років?
49. Як ШІ впливає на економічне зростання та ВВП країн, і як забезпечити справедливий розподіл переваг?
50. Які майбутні виклики етики ШІ виникнуть у контексті досягнення AGI (Artificial General Intelligence)?

11. Навчально-методичне забезпечення

Освітній процес з дисципліни «Людиноцентрований штучний інтелект» забезпечений необхідними навчально-методичними матеріалам, що розмішені в Модульному середовищі для навчання MOODLE:

1. Курс «Людиноцентрований штучний інтелект». https://msn.khmnu.edu.ua/course/view.php?id=9979

12. Матеріально-технічне та програмне забезпечення дисципліни

Необхідні інструменти включають комп'ютер, який надається для використання в лабораторіях кафедри, та стабільне інтернет-з'єднання для доступу до хмарних сервісів.

Програмне забезпечення базується на Python екосистемі з Python Software Foundation License, включаючи Python 3.8+ з pip для управління пакетами. Для розробки використовуються Jupyter Notebook або Google Colab для інтерактивної роботи. Основними бібліотеками є PyTorch з BSD ліцензією для роботи з нейронними мережами, Transformers від Hugging Face з Apache 2.0 ліцензією для великих мовних моделей, scikit-learn з BSD ліцензією для машинного навчання, а також pandas і numpy з BSD ліцензіями для обробки даних. Візуалізація здійснюється через matplotlib та seaborn.

Спеціалізовані інструменти для людиноцентрованого штучного інтелекту включають Weights & Biases з MIT ліцензією для трекінгу експериментів, fairlearn з MIT ліцензією для аналізу справедливості, LIME і SHAP з BSD ліцензіями для пояснювального штучного інтелекту, Captum з BSD ліцензією для XAI в PyTorch, та RecBole з MIT ліцензією для рекомендаційних систем.

Хмарні платформи представлені Google Colab на безкоштовному рівні для обчислень на GPU, Hugging Face Hub для безкоштовного доступу до попередньо навчених моделей, та OpenAI Gym або Gymnasium з MIT ліцензією для середовищ підкріплювального навчання.

Середовища розробки включають Visual Studio Code з розширеннями для Python під MIT ліцензією та Git для контролю версій з GPL v2 ліцензією.

Усе програмне забезпечення має відкритий код або безкоштовні академічні ліцензії, що забезпечує повну доступність для всіх студентів без додаткових витрат на ліцензування.

13. Рекомендована література:

Основна

1. A compendium of responsible artificial intelligence: / D. K. Jain, K. R. Bhatele, D. Garg, G. Dhiman. Boca Raton: CRC Press, Taylor & Francis Group, 2025. 350p.
2. Banafa A. Transformative AI: responsible, transparent, and trustworthy AI systems: Gistrup, Denmark: River Publishers, 2024. 172p.
3. Papagiannidis E., Mikalef P., Conboy K. Responsible artificial intelligence governance: A review and research framework. *The Journal of Strategic Information Systems*. 2025. Vol. 34, No. 2. URL: https://doi.org/10.1016/j.jsis.2024.101885.
4. Taherdoost H., Madanchian M., Castanho G. Balancing Innovation, Responsibility, and Ethical Consideration in AI Adoption. *Procedia Computer Science*. 2025. Vol. 258. Pp. 3284–3293. URL: https://doi.org/10.1016/j.procs.2025.04.586.
5. Madanchian M., Taherdoost H. Ethical theories, governance models, and strategic frameworks for responsible AI adoption and organizational success. *Frontiers in Artificial Intelligence*. 2025. Vol. 8. Pp. 1619029. URL: https://doi.org/10.3389/frai.2025.1619029.
6. Bayram F., Ahmed B. S. Towards Trustworthy Machine Learning in Production: An Overview of the Robustness in MLOps Approach. *ACM Computing Surveys*. 2025. Vol. 57, No. 5. Pp. 1–35. URL: https://doi.org/10.1145/3708497.
7. Amrit C., Narayanappa A. K. An analysis of the challenges in the adoption of MLOps. *Journal of Innovation & Knowledge*. 2025. Vol. 10, No. 1. Pp. 100637. URL: https://doi.org/10.1016/j.jik.2024.100637.
8. Pahune S., Akhtar Z. Transitioning from MLOps to LLMOps: Navigating the Unique Challenges of Large Language Models. *Information*. 2025. Vol. 16, No. 2. Pp. 87. URL: https://doi.org/10.3390/info16020087.
9. Ferrara E. Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies. *Sci*. 2023. Vol. 6, No. 1. Pp. 3. URL: https://doi.org/10.3390/sci6010003.
10. González-Sendino R., Serrano E., Bajo J. Mitigating bias in artificial intelligence: Fair data generation via causal models for transparent and explainable decision-making. *Future Generation Computer Systems*. 2024. Vol. 155. Pp. 384–401. URL: https://doi.org/10.1016/j.future.2024.02.023.
11. Gallegos I. O., Rossi R. A., Barrow J., Tanjim M. M., Kim S., Dernoncourt F., Yu T., Zhang R., Ahmed N. K. Bias and Fairness in Large Language Models: A Survey. *Computational Linguistics*. 2024. Vol. 50, No. 3. Pp. 1097–1179. URL:

https://doi.org/10.1162/coli_a_00524.
12. Tiribelli S., Giovanola B., Pietrini R., Frontoni E., Paolanti M. Embedding AI ethics into the design and use of computer vision technology for consumer's behaviour understanding. *Computer Vision and Image Understanding*. 2024. Vol. 248. Pp. 104142. URL: https://doi.org/10.1016/j.cviu.2024.104142.
13. Lekadir K., Frangi A. F., Porras A. R., Glocker B., Cintas C., Langlotz C. P., Weicken E. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. *BMJ (Clinical research ed.)*. 2025. Vol. 388. URL: https://doi.org/10.1136/bmj-2024-081554.
14. Alvarez J. M., Colmenarejo A. B., Elobaid A., Fabbrizzi S., Fahimi M., Ferrara A., Ghodsi S., Mougan C., Papageorgiou I., Reyero P., Russo M., Scott K. M., State L., Zhao X., Ruggieri S. Policy advice and best practices on bias and fairness in AI. *Ethics and Information Technology*. 2024. Vol. 26, No. 2. Pp. 31. URL: https://doi.org/10.1007/s10676-024-09746-w.
15. Alzubi T. M., Alzubi J. A., Singh A., Alzubi O. A., Subramanian M. A Multimodal Human-Computer Interaction for Smart Learning System. *International Journal of Human–Computer Interaction*. 2025. Vol. 41, No. 3. Pp. 1718–1728. URL: https://doi.org/10.1080/10447318.2023.2206758.
16. Tamang S., Bora D. J. Enforcement Agents: Enhancing Accountability and Resilience in Multi-Agent AI Frameworks. 2025. URL: https://doi.org/10.48550/ARXIV.2504.04070.
17. Kyriakou K., Otterbacher J. Modular oversight methodology: a framework to aid ethical alignment of algorithmic creations. *Design Science*. 2024. Vol. 10. URL: https://doi.org/10.1017/dsj.2024.23.
18. Koshiyama A., Kazim E., Treleaven P., Rai P., Szpruch L., Pavey G., Ahamat G., Leutner F., Goebel R., Knight A., Adams J., Hitrova C., Barnett J., Nachev P., Barber D., Chamorro-Premuzic T., Klemmer K., Gregorovic M., Khan S., Lomas E., Hilliard A., Chatterjee S. Towards algorithm auditing: managing legal, ethical and technological risks of AI, ML and associated algorithms. *Royal Society Open Science*. 2024. Vol. 11, No. 5. Pp. 219–234. URL: https://doi.org/10.1098/rsos.230859.
19. Billeter Y., Denzel P., Chavarriaga R., Forster O., Schilling F.-P., Brunner S., Frischknecht-Gruber C., Reif M., Weng J. MLOps as Enabler of Trustworthy AI: *2024 11th IEEE Swiss Conference on Data Science (SDS)*, Zurich, Switzerland , IEEE, May 30, 2024. Pp.37–40. URL: https://doi.org/10.1109/SDS60720.2024.00013.
20. Čop A., Bertalanič B., Fortuna C. An overview and solution for democratizing AI workflows at the network edge. *Journal of Network and Computer Applications*. 2025. Vol. 239. URL: https://doi.org/10.1016/j.jnca.2025.104180.
21. Li Z., Liang C., Peng J., Yin M. How Does the Disclosure of AI Assistance Affect the Perceptions of Writing?: *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, Miami, Florida, USA , Association for Computational Linguistics, 2024. Pp.4849–4868. URL: https://doi.org/10.18653/v1/2024.emnlp-main.279.
22. Tariq M. U. The Role of AI Ethics in Cost and Complexity Reduction: *Advances in Business Information Systems and Analytics*. 2024. Pp. 59–78. URL: https://doi.org/10.4018/979-8-3693-2643-5.ch004.
23. Kvitka A., Sosnin D., Kvitka Y., Andreieva K. The role of artificial intelligence in the development of the company's business processes. *Facta Universitatis, Series: Economics and Organization*. 2024. Pp. 47–57. URL: https://doi.org/10.22190/FUEO231206003K.
24. Gondauri D. The Impact of Artificial Intelligence on Gross Domestic Product: A Global Analysis. 2025. URL: https://doi.org/10.48550/ARXIV.2505.11989.
25. Lenharo M. ChatGPT turns two: how the AI chatbot has changed scientists' lives. *Nature*. 2024. Vol. 636, No. 8042. Pp. 281–282. URL: https://doi.org/10.1038/d41586-024-03940-y.
26. Birhane A., Kasirzadeh A., Leslie D., Wachter S. Science in the age of large language models. *Nature Reviews Physics*. 2023. Vol. 5, No. 5. Pp. 277–280. URL: https://doi.org/10.1038/s42254-023-00581-4.
27. Tunstall L., Beeching E., Lambert N., Rajani N., Huang S., Rasul K., Bartolome A., M. Rush A., Wolf T. The Alignment Handbook: URL: https://github.com/huggingface/alignment-handbook.
28. SFT Trainer: URL: https://huggingface.co/docs/trl/en/sft_trainer.
29. Google Colab: URL: https://colab.research.google.com/github/NielsRogge/Transformers-Tutorials/blob/master/Mistral/Supervised_fine_tuning_(SFT)_of_an_LLM_using_Hugging_Face_tooling.ipynb.
30. huggingface/llm_training_handbook: URL: https://github.com/huggingface/llm_training_handbook.
31. wandb/wandb: URL: https://github.com/wandb/wandb.
32. wandb/examples: URL: https://github.com/wandb/examples.
33. Weights & Biases Documentation: URL: https://docs.wandb.ai/Weights & Biases Documentation.
34. Fairlearn: URL: https://fairlearn.org.
35. Barry Becker R. K. Adult: URL: https://archive.ics.uci.edu/dataset/2.
36. Trusted-AI/AIF360: URL: https://github.com/Trusted-AI/AIF360.
37. Weerts H., Dudík M., Edgar R., Jalali A., Lutz R., Madaio M. Fairlearn: Assessing and Improving Fairness of AI Systems: URL: http://jmlr.org/papers/v24/23-0389.htmlJournal of Machine Learning Research.
38. Welcome to the SHAP documentation — SHAP latest documentation: URL: https://shap.readthedocs.io/en/latest/.
39. Local Interpretable Model-Agnostic Explanations (lime) — lime 0.1 documentation: URL: https://lime-ml.readthedocs.io/en/latest/.
40. shap/shap: URL: https://github.com/shap/shap.
41. Ribeiro M. T. C. marcotcr/lime: URL: https://github.com/marcotcr/lime.
42. Build LLM apps - Streamlit Docs: URL: https://docs.streamlit.io/.
43. Team G. Gradio Documentation: URL: https://www.gradio.app/docs.
44. Esuli A., Sebastiani F. Active Learning Strategies for Multi-Label Text Classification: *Advances in Information Retrieval*: M. Boughanem, C. Berrut, J. Mothe, C. Soule-Dupuy. Berlin, Heidelberg, Springer Berlin Heidelberg, 2009.

https://doi.org/10.1007/978-3-642-00958-7_12.
45. Streamlit Docs: URL: https://docs.streamlit.io/.
46. Amershi S., Weld D., Vorvoreanu M., Fourney A., Nushi B., Collisson P., Suh J., Iqbal S., Bennett P., Inkpen K., Teevan J., Kikin-Gil R., Horvitz E. Guidelines for Human-AI InteractionMay 01, 2019.
47. Abid A., Abdalla A., Abid A., Khan D., Alfozan A., Zou J. Gradio: Hassle-free sharing and testing of ML models in the wild: URL: https://arxiv.org/abs/1906.02569.
48. microsoft/responsible-ai-toolbox: URL: https://github.com/microsoft/responsible-ai-toolbox.
49. tensorflow/model-card-toolkit: URL: https://github.com/tensorflow/model-card-toolkit.
50. dssg/aequitas: URL: https://github.com/dssg/aequitas.
51. IBM/bias-mitigation-of-machine-learning-models-using-aif360: URL: https://github.com/IBM/bias-mitigation-of-machine-learning-models-using-aif360.
52. Partnership on AI - Home: URL: https://partnershiponai.org/Partnership on AI.
53. Zhang H. hongleizhang/RSPapers: URL: https://github.com/hongleizhang/RSPapers.
54. VectorInstitute/Recommender-Systems-Survey: URL: https://github.com/VectorInstitute/Recommender-Systems-Survey.
55. Gu Y. guyulongcs/Awesome-Deep-Learning-Papers-for-Search-Recommendation-Advertising: URL: https://github.com/guyulongcs/Awesome-Deep-Learning-Papers-for-Search-Recommendation-Advertising.
56. Yuanzi (BiYa) L. Lyz103/Recommendation-paper-daily: URL: https://github.com/Lyz103/Recommendation-paper-daily.

Додаткова

1. Raschka S. Build a Large Language Model (From Scratch). Manning Publications, 2024. 264 p.
2. Ouyang L., Wu J., Jiang X., Almeida D., Wainwright C., Mishkin P., Zhang C., Agarwal S., Slama K., Ray A., Schulman J., Hilton J., Kelton F., Miller L., Simens M., Askell A., Welinder P., Christiano P., Leike J., Lowe R. Training language models to follow instructions with human feedback. Advances in Neural Information Processing Systems. 2022. Vol. 35. P. 27730-27744.
3. Rafailov R., Sharma A., Mitchell E., Manning C. D., Ermon S., Finn C. Direct preference optimization: Your language model is secretly a reward model. Advances in Neural Information Processing Systems. 2023. Vol. 36.
4. Kreuzberger D., Kühl N., Hirschl S. Machine Learning Operations (MLOps): Overview, Definition, and Architecture. IEEE Access. 2023. Vol. 11. P. 31866-31879. DOI: 10.1109/ACCESS.2023.3262138.
5. Das Jui T., Rivas P. Fairness issues, current approaches, and challenges in machine learning models. International Journal of Machine Learning and Cybernetics. 2024. Vol. 15. P. 1-31. DOI: 10.1007/s13042-023-02083-2.
6. Caton S., Haas C. Fairness in Machine Learning: A Survey. ACM Computing Surveys. 2024. Vol. 56, No. 7. Article 166. 38 p. DOI: 10.1145/3616865.
7. Vimbi V., Shaffi N., Mahmud M. Interpreting artificial intelligence models: a systematic review on the application of LIME and SHAP in Alzheimer's disease detection. Brain Informatics. 2024. Vol. 11. P. 10. DOI: 10.1186/s40708-024-00222-1.
8. Salih A., Raisi-Estabragh Z., Galazzo I. B., Radeva P., Petersen S. E., Menegaz G., Lekadir K. A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME. Advanced Intelligent Systems. 2025. Vol. 7. P. 2400304. DOI: 10.1002/aisy.202400304.
9. Wu J., Huang Z., Hu Z., Lv C. Toward Human-in-the-loop AI: Enhancing Deep Reinforcement Learning via Real-time Human Guidance for Autonomous Driving. Engineering. 2023. Vol. 21, No. 2. P. 75-91. DOI: 10.1016/j.eng.2022.05.017.
10. Amershi S., Weld D., Vorvoreanu M., Fourney A., Nushi B., Collisson P., Suh J., Iqbal S., Bennett P., Inkpen K., Teevan J., Kikin-Gil R., Horvitz E. Guidelines for Human-AI Interaction. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019. P. 1-13. DOI: 10.1145/3290605.3300233.
11. Esuli A., Sebastiani F. Active Learning Strategies for Multi-Label Text Classification. Advances in Information Retrieval / M. Boughanem, C. Berrut, J. Mothe, C. Soule-Dupuy. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. P. 102-117. DOI: 10.1007/978-3-642-00958-7_12.
12. Deldjoo Y., Jannach D., Bellogin A., Difonzo A., Zanzonelli D. Fairness in recommender systems: research landscape and future directions. User Modeling and User-Adapted Interaction. 2023. Vol. 34, No. 1. P. 59–108. DOI: 10.1007/s11257-023-09364-z.
13. Dige O., Arneja D., Yau T. F., Zhang Q., Bolandraftar M., Zhu X., Khattak F. K. Can Machine Unlearning Reduce Social Bias in Language Models? // Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track. 2024. P. 954–969.
14. Gallegos I. O., Rossi R. A., Barrow J., Tanjim M. M., Kim S., Dernoncourt F. et al. Bias and Fairness in Large Language Models: A Survey // Computational Linguistics. 2024. Vol. 50, No. 3. P. 1097–1179.Li J., Tang Z., Liu X., Spirtes P., Zhang K., Leqi L., Liu Y. Steering LLMs towards Unbiased Responses: A Causality-Guided Debiasing Framework. arXiv preprint arXiv:2403.08743. 2024.
15. Lin Z., Guan S., Zhang W., Zhang H., Li Y., Zhang H. Towards trustworthy LLMs: a review on debiasing and dehallucinating in large language models. Artificial Intelligence Review. 2024. Vol. 57, No. 9. Pp. 243. URL: https://doi.org/10.1007/s10462-024-10896-y.
16. Huang D., Zhang J. M., Bu Q., Xie X., Chen J., Cui H. Bias Testing and Mitigation in LLM-based Code Generation. ACM Transactions on Software Engineering and Methodology. 2025. Pp. 3724117. URL: https://doi.org/10.1145/3724117.

14. Інформаційні ресурси

1. Модульне середовище для навчання. URL: https://msn.khmnu.edu.ua/course/view.php?id=9979
2. Електронна бібліотека ХНУ. URL: http://library.khmnu.edu.ua
3. Інституційний репозитарій ХНУ. URL: https://elar.khmnu.edu.ua/home

ЛЮДИНОЦЕНТРОВАНИЙ ШТУЧНИЙ ІНТЕЛЕКТ

Тип дисципліни	Обов'язкова
Рівень вищої освіти	Третій (доктор філософії)
Мова викладання	Українська
Семестр	1
Кількість встановлених кредитів ЄКТС	5
Форми навчання, для яких викладається дисципліна	Очна (денна)

Результати навчання. Після вивчення дисципліни студент повинен: застосовувати принципи відповідального ШІ та етичні стандарти у розробці інтелектуальних систем; інтегрувати MLOps процеси з принципами етичного ШІ у хмарних середовищах; виявляти та мітигувати різні типи упередженості на всіх етапах досліджень; розробляти та впроваджувати методи пояснюваності та інтерпретації ШІ моделей; проєктувати людиноцентровані інтерфейси з урахуванням психологічних аспектів взаємодії; проводити аудит та сертифікацію ШІ систем відповідно до регулятивних вимог; реалізовувати гармонізацію великих мовних моделей методами SFT; розробляти етично обумовлені рекомендаційні системи; застосовувати Human-in-the-Loop підходи для автономних систем; операціоналізувати етичні принципи у практичні бізнес-рішення з використанням ШІ технологій.

Зміст навчальної дисципліни. Теоретичні основи людиноцентрованого штучного інтелекту; гармонізація великих мовних моделей методом SFT; етапи розробки людиноцентрованої моделі машинного навчання з використанням W&B; виявлення та зменшення упередженості в моделях ШІ; розробка XAI методів з використанням NormEnsembleXAI; Human-in-the-Loop машинне навчання для автономних систем; Парето-оптимальна оцінка справедливості в рекомендаційних системах; мітігація упереджень у великих мовних моделях для етично обумовлених рекомендаційних систем.

Пререквізити – вихідна.

Кореквізити – моделювання та інтелектуальна обробка даних (ОФП.03), педагогічна (викладацька) практика (ОФП.05).

Запланована навчальна діяльність: Мінімальний обсяг навчальних занять в одному кредиті ЄКТС навчальної дисципліни для ***першого*** (бакалаврського) рівня вищої освіти за денною формою здобуття освіти становить 8 годин на 1 кредит ЄКТС.

Форми (методи) навчання: лекції (з використанням методів візуалізації, проблемного навчання, мотиваційних прийомів, інформаційно-комунікаційних технологій); лабораторні заняття: з використанням методів комп'ютерного моделювання, аналіз проблемних ситуацій, пояснення, дискусія тощо, самостійна робота: робота над засвоєнням теоретичного матеріалу, виконання індивідуальних та завдань, підготовка до поточного та підсумкового контролю тощо) з використанням інформаційно-комп'ютерних технологій та технологій дистанційного навчання.

Форми оцінювання результатів навчання: захист лабораторних робіт, виконання індивідуального задання.

Вид семестрового контролю: іспит.

Навчальні ресурси:

1. A compendium of responsible artificial intelligence: / D. K. Jain, K. R. Bhatele, D. Garg, G. Dhiman. Boca Raton: CRC Press, Taylor & Francis Group, 2025. 350p.
2. Banafa A. Transformative AI: responsible, transparent, and trustworthy AI systems: Gistrup, Denmark: River Publishers, 2024. 172p.
3. Papagiannidis E., Mikalef P., Conboy K. Responsible artificial intelligence governance: A review and research framework. *The Journal of Strategic Information Systems*. 2025. Vol. 34, No. 2. URL: https://doi.org/10.1016/j.jsis.2024.101885.
4. Taherdoost H., Madanchian M., Castanho G. Balancing Innovation, Responsibility, and Ethical Consideration in AI Adoption. *Procedia Computer Science*. 2025. Vol. 258. Pp. 3284–3293. URL: https://doi.org/10.1016/j.procs.2025.04.586.
5. Madanchian M., Taherdoost H. Ethical theories, governance models, and strategic frameworks for responsible AI adoption and organizational success. *Frontiers in Artificial Intelligence*. 2025. Vol. 8. Pp. 1619029. URL: https://doi.org/10.3389/frai.2025.1619029.
6. Bayram F., Ahmed B. S. Towards Trustworthy Machine Learning in Production: An Overview of the Robustness in MLOps Approach. *ACM Computing Surveys*. 2025. Vol. 57, No. 5. Pp. 1–35. URL: https://doi.org/10.1145/3708497.
7. Amrit C., Narayanappa A. K. An analysis of the challenges in the adoption of MLOps. *Journal of Innovation & Knowledge*. 2025. Vol. 10, No. 1. Pp. 100637. URL: https://doi.org/10.1016/j.jik.2024.100637.
8. Pahune S., Akhtar Z. Transitioning from MLOps to LLMOps: Navigating the Unique Challenges of Large Language Models. *Information*. 2025. Vol. 16, No. 2. Pp. 87. URL: https://doi.org/10.3390/info16020087.
9. Ferrara E. Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies. *Sci*. 2023. Vol. 6, No. 1. Pp. 3. URL: https://doi.org/10.3390/sci6010003.
10. González-Sendino R., Serrano E., Bajo J. Mitigating bias in artificial intelligence: Fair data generation via causal

models for transparent and explainable decision-making. *Future Generation Computer Systems*. 2024. Vol. 155. Pp. 384–401. URL: https://doi.org/10.1016/j.future.2024.02.023.
11. Gallegos I. O., Rossi R. A., Barrow J., Tanjim M. M., Kim S., Dernoncourt F., Yu T., Zhang R., Ahmed N. K. Bias and Fairness in Large Language Models: A Survey. *Computational Linguistics*. 2024. Vol. 50, No. 3. Pp. 1097–1179. URL: https://doi.org/10.1162/coli_a_00524.
12. Tiribelli S., Giovanola B., Pietrini R., Frontoni E., Paolanti M. Embedding AI ethics into the design and use of computer vision technology for consumer's behaviour understanding. *Computer Vision and Image Understanding*. 2024. Vol. 248. Pp. 104142. URL: https://doi.org/10.1016/j.cviu.2024.104142.
13. Lekadir K., Frangi A. F., Porras A. R., Glocker B., Cintas C., Langlotz C. P., Weicken E. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. *BMJ (Clinical research ed.)*. 2025. Vol. 388. URL: https://doi.org/10.1136/bmj-2024-081554.
14. Alvarez J. M., Colmenarejo A. B., Elobaid A., Fabbrizzi S., Fahimi M., Ferrara A., Ghodsi S., Mougan C., Papageorgiou I., Reyero P., Russo M., Scott K. M., State L., Zhao X., Ruggieri S. Policy advice and best practices on bias and fairness in AI. *Ethics and Information Technology*. 2024. Vol. 26, No. 2. Pp. 31. URL: https://doi.org/10.1007/s10676-024-09746-w.
15. Alzubi T. M., Alzubi J. A., Singh A., Alzubi O. A., Subramanian M. A Multimodal Human-Computer Interaction for Smart Learning System. *International Journal of Human–Computer Interaction*. 2025. Vol. 41, No. 3. Pp. 1718–1728. URL: https://doi.org/10.1080/10447318.2023.2206758.
16. Tamang S., Bora D. J. Enforcement Agents: Enhancing Accountability and Resilience in Multi-Agent AI Frameworks. 2025. URL: https://doi.org/10.48550/ARXIV.2504.04070.
17. Kyriakou K., Otterbacher J. Modular oversight methodology: a framework to aid ethical alignment of algorithmic creations. *Design Science*. 2024. Vol. 10. URL: https://doi.org/10.1017/dsj.2024.23.
18. Koshiyama A., Kazim E., Treleaven P., Rai P., Szpruch L., Pavey G., Ahamat G., Leutner F., Goebel R., Knight A., Adams J., Hitrova C., Barnett J., Nachev P., Barber D., Chamorro-Premuzic T., Klemmer K., Gregorovic M., Khan S., Lomas E., Hilliard A., Chatterjee S. Towards algorithm auditing: managing legal, ethical and technological risks of AI, ML and associated algorithms. *Royal Society Open Science*. 2024. Vol. 11, No. 5. Pp. 219–234. URL: https://doi.org/10.1098/rsos.230859.
19. Billeter Y., Denzel P., Chavarriaga R., Forster O., Schilling F.-P., Brunner S., Frischknecht-Gruber C., Reif M., Weng J. MLOps as Enabler of Trustworthy AI: *2024 11th IEEE Swiss Conference on Data Science (SDS)*, Zurich, Switzerland , IEEE, May 30, 2024. Pp.37–40. URL: https://doi.org/10.1109/SDS60720.2024.00013.
20. Čop A., Bertalanič B., Fortuna C. An overview and solution for democratizing AI workflows at the network edge. *Journal of Network and Computer Applications*. 2025. Vol. 239. URL: https://doi.org/10.1016/j.jnca.2025.104180.
21. Li Z., Liang C., Peng J., Yin M. How Does the Disclosure of AI Assistance Affect the Perceptions of Writing?: *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, Miami, Florida, USA , Association for Computational Linguistics, 2024. Pp.4849–4868. URL: https://doi.org/10.18653/v1/2024.emnlp-main.279.
22. Tariq M. U. The Role of AI Ethics in Cost and Complexity Reduction: *Advances in Business Information Systems and Analytics*. 2024. Pp. 59–78. URL: https://doi.org/10.4018/979-8-3693-2643-5.ch004.
23. Kvitka A., Sosnin D., Kvitka Y., Andreieva K. The role of artificial intelligence in the development of the company's business processes. *Facta Universitatis, Series: Economics and Organization*. 2024. Pp. 47–57. URL: https://doi.org/10.22190/FUEO231206003K.
24. Gondauri D. The Impact of Artificial Intelligence on Gross Domestic Product: A Global Analysis. 2025. URL: https://doi.org/10.48550/ARXIV.2505.11989.
25. Lenharo M. ChatGPT turns two: how the AI chatbot has changed scientists' lives. *Nature*. 2024. Vol. 636, No. 8042. Pp. 281–282. URL: https://doi.org/10.1038/d41586-024-03940-y.
26. Birhane A., Kasirzadeh A., Leslie D., Wachter S. Science in the age of large language models. *Nature Reviews Physics*. 2023. Vol. 5, No. 5. Pp. 277–280. URL: https://doi.org/10.1038/s42254-023-00581-4.
27. Tunstall L., Beeching E., Lambert N., Rajani N., Huang S., Rasul K., Bartolome A., M. Rush A., Wolf T. The Alignment Handbook: URL: https://github.com/huggingface/alignment-handbook.
28. SFT Trainer: URL: https://huggingface.co/docs/trl/en/sft_trainer.
29. Google Colab: URL: https://colab.research.google.com/github/NielsRogge/Transformers-Tutorials/blob/master/Mistral/Supervised_fine_tuning_(SFT)_of_an_LLM_using_Hugging_Face_tooling.ipynb.
30. huggingface/llm_training_handbook: URL: https://github.com/huggingface/llm_training_handbook.
31. wandb/wandb: URL: https://github.com/wandb/wandb.
32. wandb/examples: URL: https://github.com/wandb/examples.
33. Weights & Biases Documentation: URL: https://docs.wandb.ai/Weights & Biases Documentation.
34. Fairlearn: URL: https://fairlearn.org.
35. Barry Becker R. K. Adult: URL: https://archive.ics.uci.edu/dataset/2.
36. Trusted-AI/AIF360: URL: https://github.com/Trusted-AI/AIF360.
37. Weerts H., Dudík M., Edgar R., Jalali A., Lutz R., Madaio M. Fairlearn: Assessing and Improving Fairness of AI Systems: URL: http://jmlr.org/papers/v24/23-0389.htmlJournal of Machine Learning Research.
38. Welcome to the SHAP documentation — SHAP latest documentation: URL: https://shap.readthedocs.io/en/latest/.
39. Local Interpretable Model-Agnostic Explanations (lime) — lime 0.1 documentation: URL: https://lime-ml.readthedocs.io/en/latest/.
40. shap/shap: URL: https://github.com/shap/shap.
41. Ribeiro M. T. C. marcotcr/lime: URL: https://github.com/marcotcr/lime.

42. Build LLM apps - Streamlit Docs: URL: https://docs.streamlit.io/.
43. Team G. Gradio Documentation: URL: https://www.gradio.app/docs.
44. Esuli A., Sebastiani F. Active Learning Strategies for Multi-Label Text Classification: *Advances in Information Retrieval*: M. Boughanem, C. Berrut, J. Mothe, C. Soule-Dupuy. Berlin, Heidelberg, Springer Berlin Heidelberg, 2009. https://doi.org/10.1007/978-3-642-00958-7_12.
45. Streamlit Docs: URL: https://docs.streamlit.io/.
46. Amershi S., Weld D., Vorvoreanu M., Fourney A., Nushi B., Collisson P., Suh J., Iqbal S., Bennett P., Inkpen K., Teevan J., Kikin-Gil R., Horvitz E. Guidelines for Human-AI InteractionMay 01, 2019.
47. Abid A., Abdalla A., Abid A., Khan D., Alfozan A., Zou J. Gradio: Hassle-free sharing and testing of ML models in the wild: URL: https://arxiv.org/abs/1906.02569.
48. microsoft/responsible-ai-toolbox: URL: https://github.com/microsoft/responsible-ai-toolbox.
49. tensorflow/model-card-toolkit: URL: https://github.com/tensorflow/model-card-toolkit.
50. dssg/aequitas: URL: https://github.com/dssg/aequitas.
51. IBM/bias-mitigation-of-machine-learning-models-using-aif360: URL: https://github.com/IBM/bias-mitigation-of-machine-learning-models-using-aif360.
52. Partnership on AI - Home: URL: https://partnershiponai.org/Partnership on AI.
53. Zhang H. hongleizhang/RSPapers: URL: https://github.com/hongleizhang/RSPapers.
54. VectorInstitute/Recommender-Systems-Survey: URL: https://github.com/VectorInstitute/Recommender-Systems-Survey.
55. Gu Y. guyulongcs/Awesome-Deep-Learning-Papers-for-Search-Recommendation-Advertising: URL: https://github.com/guyulongcs/Awesome-Deep-Learning-Papers-for-Search-Recommendation-Advertising.
56. Yuanzi (BiYa) L. Lyz103/Recommendation-paper-daily: URL: https://github.com/Lyz103/Recommendation-paper-daily.
57. Модульне середовище для навчання. URL: Модульне середовище для навчання. URL: https://msn.khmnu.edu.ua/course/view.php?id=9979
58. Електронна бібліотека університету. URL: http://library.khmnu.edu.ua

Викладач: д-р. техн. н., проф., Едуард МАНЗЮК