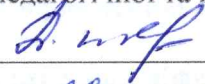


**Приватне акціонерне товариство
«Приватний вищий навчальний заклад
«Запорізький інститут економіки та інформаційних технологій»**

Кафедра інформаційних технологій

“ЗАТВЕРДЖУЮ”

Проректор з науково-педагогічної та навчальної роботи

 Дмитро ШВЕЦЬ
“ 28 ” 08 2026 року

**СИЛАБУС ОСВІТНЬОГО КОМПОНЕНТА
ОК 07 СУЧАСНІ ІНТЕЛЕКТУАЛЬНІ ТЕХНОЛОГІЇ**

Освітня програма: **КОМП'ЮТЕРНА ІНЖЕНЕРІЯ**

Спеціальності: **F7 КОМП'ЮТЕРНА ІНЖЕНЕРІЯ**

Рівень вищої освіти: **другий (магістерський)**

2026-2027 н.р.

Мова викладання: українська

Прізвище, ім'я та по-батькові викладача(ів)/розробника(ів):

Д.т.н., доцент Котов Ігор Анатолійович

Електронна адреса викладача(ів): і. kotov@econom.zp.ua

Дні занять за розкладом _____ *згідно з розкладом занять*

Графік консультацій (онлайн консультації) _____ *згідно з графіком консультацій*

Сторінка курсу в Moodle: <https://moodle.zieit.zp.ua/course/view.php?id=1198> _____

Силабус схвалено на засіданні кафедри: інформаційних технологій

Протокол від “28” серпня 2026р. № 1

Завідувач кафедри інформаційних технологій

 (Левицький С.І.)

Силабус погоджено

Начальник навчального відділу

 Ольга Сташкевічус

1. Опис освітнього компонента

Анотація освітнього компонента (Призначення освітнього компонента)	Призначення освітнього компонента — формування у здобувачів вищої освіти поглиблених теоретичних знань та практичних компетентностей у сфері інженерного супроводу, розгортання, оптимізації, моніторингу та підтримки життєвого циклу сучасних інтелектуальних систем. Курс орієнтований на підготовку фахівців, здатних забезпечувати надійне функціонування ШІ-рішень у продакшен-середовищі, керувати обчислювальними ресурсами нейромереж, організовувати наскрізну інфраструктуру та вирішувати складні системно-технічні задачі у межах експлуатації інтелектуальних платформ.
Мета вивчення	Формування у магістрів спеціальності «Комп'ютерна інженерія» комплексу професійних знань і практичних навичок, необхідних для системного дизайну інфраструктури ШІ, розгортання та адміністрування нейромережових моделей, забезпечення їхньої відмовостійкості, масштабованості та ефективного моніторингу під час експлуатації.
Завдання освітнього компонента	У результаті опанування освітнього компонента здобувач вищої освіти повинен: Знати: Архітектурні засади побудови та розгортання інфраструктури для інтелектуальних систем. Особливості апаратно-програмного забезпечення для виконання та прискорення нейронних мереж. Принципи представлення нейромереж як обчислювальних графів для їх подальшої оптимізації під час виконання в інфраструктурі. Методології наскрізного дизайну ШІ-систем та аналізу інженерних компромісів (квадрат компромісів: продуктивність, ресурсомісткість, вартість, надійність). Розуміти: Роль інженера у підтримці та життєвому циклі інтелектуальних систем (MLOps / SysOps). Вплив архітектури обчислювальних графів на навантаження на апаратні ресурси (CPU/GPU/TPU) під час роботи ШІ-систем. Проблематику деградації моделей, виникнення збоїв у продакшені та методи їх діагностики. Логіку прийняття інженерних рішень в умовах обмеженості ресурсів (квадрат компромісів). Застосовувати: Інструменти розгортання, контейнеризації та оркестрації для підтримки роботи інтелектуальних систем. Методи моніторингу, логування та трасування стану нейромереж і супутніх сервісів у реальному часі. Підходи до оптимізації виконання моделей на інфраструктурному рівні. Практики наскрізного супроводу ШІ-інфраструктури для забезпечення стабільності, безпеки та високої продуктивності сервісів.
Передумови для вивчення освітнього компонента (пререквізити)	Знання та навички, отримані під час навчання та успішного завершення попереднього рівня освіти. Базові знання архітектури комп'ютерів і комп'ютерних мереж, операційних систем, базове розуміння принципів роботи сучасних програмних інтерфейсів та баз даних, основи програмування.

Загальні та спеціальні компетентності	<i>ЗК3.Здатність проводити дослідження на відповідному рівні.</i> <i>ЗК4.Здатність до пошуку, оброблення та аналізу інформації з різних джерел.</i> <i>ЗК6Здатність виявляти, ставити та вирішувати проблеми.</i> <i>СК2. Здатність розробляти алгоритмічне та програмне забезпечення, компоненти комп'ютерних систем та мереж, Інтернет додатків, кіберфізичних систем з використанням сучасних методів і мов програмування, а також засобів і систем автоматизації проектування.</i> <i>СК3. Здатність проектувати комп'ютерні системи та мережі з урахуванням цілей, обмежень, технічних, економічних та правових аспектів.</i> <i>СК6. Здатність використовувати та впроваджувати нові технології, включаючи технології розумних, мобільних, зелених і безпечних обчислень, брати участь в модернізації та реконструкції комп'ютерних систем та мереж, різноманітних вбудованих і розподілених додатків, зокрема з метою підвищення їх ефективності.</i>
Результати навчання	<i>РН3. Будувати та досліджувати моделі комп'ютерних систем і мереж, оцінювати їх адекватність, визначати межі застосовності.</i> <i>РН4. Застосовувати спеціалізовані концептуальні знання, що включають сучасні наукові здобутки у сфері комп'ютерної інженерії, необхідні для професійної діяльності, оригінального мислення та проведення досліджень, критичного осмислення проблем інформаційних технологій та на межі галузей знань.</i> <i>РН8. Застосовувати знання технічних характеристик, конструктивних особливостей, призначення і правил експлуатації програмно-технічних засобів комп'ютерних систем та мереж для вирішення складних задач комп'ютерної інженерії та дотичних проблем.</i> <i>РН9. Розробляти програмне забезпечення для вбудованих і розподілених застосувань, мобільних і гібридних систем.</i> <i>РН15. Обирати та використовувати ефективні методи та інструментальні засоби моделювання комп'ютерних систем та мереж для забезпечення оптимальних параметрів систем, що проєктуються</i>

2. Обсяг та ознаки освітнього компонента

Найменування показників	Галузь знань, напрям підготовки, освітньо-кваліфікаційний рівень	Характеристика освітнього компонента		
		денна форма навчання	заочна та дистанційна форма навчання	
Кількість кредитів – 3	Галузі знань: <i>Галузь знань: F«Інформаційні технології»</i>	Обов’язковий <i>(згідно НП)</i>		
Модулів -1	Спеціальність: <i>F7«Комп’ютерна інженерія»</i>			
Змістових модулів – 4		Рік підготовки <i>(всі дані згідно РП)</i>		
Загальна кількість годин – 90		1-й		1-й
		Семестр		
	1.-й	1.-й		
	Лекції			
	10 год	4 год		
	Практичні, семінарські			
	20 год	6 год		
	Лабораторні			
	Самостійна робота			
	60 год.	80 год.		
В т.ч. індивідуальні завдання:				
Вид контролю*:				
Залік		Залік		

3. Дидактична карта освітнього компонента

Назви змістовних модулів і тем	Кількість годин											
	денна форма						заочна форма					
	усього	у тому числі					усього	у тому числі				
		лек	прак	лаб	сем	с.р.		лек	прак	лаб	сем	с.р.
1	2	3	4	5	6	7	8	9	10	11	12	13
Модуль 1												
Змістовний модуль 1. Архітектура та системний дизайн інтелектуальних систем												
Тема 1. <i>Основи інтелектуальних систем (Введення в ІІІ). Нейронні мережі. Нейромережа як обчислювальний граф. Наскрізний дизайн ІІІ-систем та Квадрат компромісів</i>	30	4	6			20	30	1	2			27
Змістовний модуль 2. Проектування, стиснення та компіляція моделей												
Тема 2 <i>Архітектурні шаблони сучасного Deep Learning. Інженерія стиснення моделей (Model Co-design). Графові компілятори ІІІ.</i>	20	2	6			12	20	1	2			17
Змістовний модуль 3. Клієнтські обчислення: Edge AI, Web-AI та Mobile Systems												
Тема 3. <i>Інтелектуальні веб- та мобільні додатки. Вбудовані та IoT системи (TinyML). Архітектура розподілених даних.</i>	20	2	4			14	20	1	1			18
Змістовний модуль 4. Платформи ІІІ, MLOps та LLM-інженерія												
Тема 4. <i>Конвеєри MLOps (Machine Learning Operations). Масштабування та хмарна інфраструктура. . LLMOps та розробка систем на базі LLM.</i>	20	2	4			14	20	1	1			18
Разом за 1 модулем	90	10	20			60	90	4	6			80
Усього годин	90	10	20			60	90	4	6			80

3.1. Карта лекційних занять

№ теми	Теми лекційних занять	План теми	Годин для	
			денна	заочна
1	Тема 1. Основи інтелектуальних систем Нейронні мережі. Нейромережа як обчислювальний граф. Наскрізний дизайн ШІ-систем та Квадрат компромісів	Визначення та фундаментальні поняття Еволюційна класифікація ШІ (за рівнем автономності та можливостей) Таксономія та структура галузі ШІ. Парадигма обчислень: Класичний софт vs Машичне навчання Життєвий цикл ШІ-проект. Навчання як оптимізація функцій (Градiєнтний спуск). Область застосування нейромережевої моделі. Біологічний нейрон та формальна модель нейрона Маккалокі та Піттса. Перцептрон. Алгоритми навчання нейронної мережі. Квадрат компромісів. Глибокий аналіз метрик на базі структури нейромереж. Енергоспоживання та "Пляшкове горло" пам'яті	4	1
2	Тема 2. Архітектурні шаблони сучасного Deep Learning. Інженерія стиснення моделей. Графові компілятори ШІ.	Deep Learning з інженерної точки зору. Патерн Transformers. Патерн Graph Neural Networks. Патерн Diffusion Models. Архітектурні модифікації для подолання фізичних обмежень заліза. Сутність сучасного підходу до балансу квадрату компромісів. Квантування. Структурний та неструктурний прунінг. Дистиляція знань. Чому звичайні компілятори не підходять для ШІ? Операторне злиття. Дворівнева архітектура компілятора. Основні сучасні графові компілятори для ШІ.	2	1
3	Тема 3. Інтелектуальні веб- та мобільні додатки. Вбудовані та IoT системи (TinyML). Архітектура розподілених даних.	Загальні підходи. WebGPU. ONNX Runtime Web. TensorFlow.js. CoreML на iOS. NNAPI на Android Загальні підходи та обмеження hardware. Системна оптимізація пам'яті: статичне виділення тензорів. TensorFlow Lite for Microcontrollers. Патерни Edge-to-Cloud обчислень. Обробка потокових даних з датчиків та камер. Концепція федеративного навчання (Federated Learning).	2	1
4	Тема 4. Конвеєри MLOps (Machine Learning Operations). Масштабування та хмарна інфраструктура. LLMOps та розробка систем на базі LLM.	Концепція MLOps та методологія керування життєвим циклом моделей. Версіонування даних та моделей за допомогою DVC. Автоматизація життєвого циклу за допомогою MLflow та ClearML. Моніторинг деградації моделей: Data Drift та Concept Drift. Архітектурні виклики промислового сервінгу моделей. Технології високопродуктивного сервінгу. Оркестрація інтелектуальних мікросервісів Специфіка парадигми LLMOps та виклики інженерії мовних моделей. Патерн RAG (Retrieval-Augmented Generation) та векторні бази даних	2	1
Разом			10	4

3.2. Карта практичних (лабораторних, семінарських) занять

№ теми	Назва теми	Назва роботи	Годин для	
			денна	заочна
3	<i>Тема 1. Основи інтелектуальних систем Нейронні мережі. Нейромережа як обчислювальний граф. Наскрізний дизайн ШІ-систем та Квадрат компромісів</i>	Практична робота №1. Парадигма обчислювальних графів, високорівневі абстракції шарів та інженерні компроміси в PyTorch	6	2
4	<i>Тема 2. Архітектурні шаблони сучасного Deep Learning. Інженерія стиснення моделей. Графові компілятори ШІ.</i>	Практична робота №2. Оптимізація обчислювальних графів: Від архітектурних патернів до стиснення моделей	6	2
5	<i>Тема 3. Інтелектуальні веб- та мобільні додатки. Вбудовані та IoT системи (TinyML). Архітектура розподілених даних.</i>	Практична робота №3. Клієнтські та розподілені обчислення	4	1
6	<i>Тема 4. Конвеєри MLOps (Machine Learning Operations). Масштабування та хмарна інфраструктура. . LLMOps та розробка систем на базі LLM.</i>	Практична робота №4. Розгортання RAG-конвеєрів та аудит метрик LLMOps	4	1
Разом			20	6

3.3. Карта самостійної роботи

№з/п	Назва теми	Вид роботи	Кількість годин	
			д	з
1	<i>Тема 1. Основи інтелектуальних систем Нейронні мережі. Нейромережа як обчислювальний граф. Наскрізний дизайн ШІ-систем та Квадрат компромісів</i>	Опрацювання конспекту з тем 1.1, 1.2, 1.3, 1.4. Тестування до теми 1	20	27
2	<i>Тема 2. Архітектурні шаблони сучасного Deep Learning. Інженерія стиснення моделей. Графові компілятори ШІ.</i>	Опрацювання конспекту з теми 2.1, 2.2, 2.3 Тестування до теми 2	12	17
3	<i>Тема 3. Інтелектуальні веб- та мобільні додатки. Вбудовані та IoT системи (TinyML). Архітектура розподілених даних.</i>	Опрацювання конспекту з теми 3.1, 3.2, 3.3 Тестування до теми 3	14	18
4	<i>Тема 4. Конвеєри MLOps (Machine Learning Operations). Масштабування та хмарна інфраструктура. . LLMOps та розробка систем на базі LLM.</i>	Опрацювання конспекту з теми 4.1,4.2,4.3. Тестування до теми 4	14	18
Разом			60	80

4. Індивідуальні завдання

Необхідно підготувати та захистити індивідуальне завдання (наведені на сторінці курс у складі практичних робіт).

5. Методи навчання	<i>Дидактична карта даної дисципліни</i> передбачає проведення лекційних та практичних занять, самостійної роботи. Основна мета лекційних занять — формування орієнтувальної теоретичної основи з сучасних інтелектуальних систем та технологій для подальшого свідомого засвоєння студентами навчального матеріалу дисципліни. Лекційні заняття з даної дисципліни будуються із застосуванням графічно-символьних методів. Лекційний курс проводиться згідно з планом у відповідності до даного силабусу дисципліни з метою розкриття теоретичних і практичних засад оптимізації підтримки використання інтелектуальних технологій з боку комп'ютерного заліза для набуття знань та умінь у відповідності до освітньо-професійної програми підготовки магістрів. Практичні заняття проводяться у формі практикуму в спеціально обладнаних комп'ютерних аудиторіях із застосуванням сучасних хмарних середовищ розробки та засобів віртуалізації. Основна мета практичних робіт — набуття студентами практичних навичок з дослідження сучасних підходів до реалізації складних інтелектуальних систем, а також систематизація й узагальнення здобутих під час лекційних занять теоретичних знань. Практичні роботи виконуються згідно з методичними вказівками з урахуванням самостійних завдань, що регламентуються викладачем. Результати виконаних завдань оформлюються в індивідуальний звіт про виконану роботу. Кожен студент повинен самостійно виконати всі практичні роботи, захистити їх та оформити підсумкові результати. <i>Самостійна робота студента</i> -опрацювання теоретичного матеріалу, винесеного на самостійну роботу через вивчення відповідних лекцій та проходження проміжних тестів. - підготовка до підсумкового контролю. Кожен студент має доступ до самостійної роботи у комп'ютерних аудиторіях інформаційно-обчислювального центру. <i>Консультації з дисципліни</i> проводяться згідно з розкладом проведення консультацій, наданого кафедрою та затвердженого навчальним відділом. Консультації проводяться на платформі ЦДО, в аудиторії (закріпленої за кафедрою або викладачем) або за допомогою відео-конференції. Консультаційна робота спрямована на допомогу в опануванні студентами методології певної теми чи розділу, методами самостійної навчальної роботи 1. Під час проведення лекцій використовуються словесні (пояснення, розповідь, лекція, бесіда), наочні (ілюстрація, демонстрація), практичні вправи. 2. Логічні - індуктивні (від одиничного до загального) та дедуктивні (від загального окремого) умовиводи використовуються під час проведення практичних занять 3. Проблемно-пошукові, дослідницькі, евристичні, аналітичні методи та форми роботи використовуються під час самостійної роботи студентів та практичних заняттях.
6. Система контролю та	Система складається з поточного, модульного та підсумкового контролю:

оцінювання.	<u>Поточний контроль</u> полягає у тому, що здобувачі освіти виконують практичні роботи з кожного розділу у відповідні аудиторні часи та у часи самостійної роботи та проходять тести. Практична частина, як форма поточного контролю, складається з виконання запропонованого прикладу та самостійного розв'язування додаткових завдань Виконана робота комплексно оцінюється викладачем, враховуючи такі критерії: 1. правильність одержаних результатів практичної роботи; 2. застосування раціонального методу рішення задач; 3. наявність звіту з практичної роботи (форма якого запропонована у завданні до кожної роботи на сторінці курсу) Результат виконання кожної практичної роботи оцінюється окремо. Крім того, студенти мають пройти комп'ютерний тест за матеріалами кожної лекції Самостійна робота включає в себе теоретичне вивчення питань, що узагальнюють всі розділи курсу та формують системний погляд на проблематику. Їхній максимальний сумарний рейтинговий бал становить 60 балів. <i>Модульний контроль</i> проводиться у відповідності до наказу ректора «Про проведення поточного та семестрового контролю успішності студентів інституту» згідно за розкладом, встановленим навчальним відділом. Модульний контроль проводиться на відповідному до графіку освітнього процесу тижні. Формою проведення модульного контролю є виконання модульної контрольної роботи у середовищі moodle у формі тестів. Тривалість складання модульної контрольної роботи не перевищує двох академічних годин. Максимальний рейтинговий бал при цьому не перевищує 40 балів. <i>Підсумковим контролем</i> засвоєння дисципліни є екзамен. У відповідності до стандарту підприємства підсумковий контроль автоматично проставляється як сума рейтингових балів поточного та модульного контролю. При наявності мінімально необхідної кількості балів поточного контролю (від 20 балів) студент має змогу скласти екзамен(залік) під час сесії, згідно за розкладом, встановленим навчальним відділом. Кількість максимальних рейтингових балів становить 40. Ці бали замінюють рейтингові бали модульного контролю, та у разі достатньої кількості з поточним контролем ставиться відповідна оцінка.
7. Технічне й програмне забезпечення/обладнання.	З метою якісної підготовки фахівців кожному студенту на базі ОЦ ЗІЕІТ надається спеціально сформований та схвалений рішенням кафедри програмно-апаратний комплекс навчально-методичного забезпечення, що містить наступне: - обліковий запис (профіль користувача) на сервері обчислювального центру, що забезпечує санкціонований вхід до комплексу, регламентує певні права та правила користування, надає можливість контролю над відвідуванням; - персональний комп'ютер з мережним постійним запам'ятовуючим пристроєм, що підключається; - повний комплект системного та прикладного програмного забезпечення (вільного користування або з наявністю відповідної ліцензії); - комплект навчально-методичного матеріалу, що містить у повному обсязі методичне забезпечення, базову та допоміжну літературу.

	Для повноцінного вивчення дисципліни надається навчально-методичний матеріал, що міститься у профілі користувача у системі moodle. Для доступу до якого необхідно мати персональний комп'ютер (ноутбук, мобільний пристрій (телефон, планшет) зі стабільним інтернет-зв'язком Практичні роботи з курсу не передбачають встановлення спеціального програмного забезпечення та виконуються в хмарному середовищі Google Colab
8. Неформальна/інформальна освіта	Здобувач освіти має можливість перезарахувати окремі теми (відповіді на семінарах, практичні завдання) або курс шляхом: проходження професійних курсів чи тренінгів, онлайн-освіти, професійних стажувань, у сфері, що відповідає навчальним цілям дисципліни. При цьому контрольні заходи з дисципліни виконуються на загальних підставах. Для зарахування необхідно надати: сертифікат (електронний або друкований) про проходження курсу/стажування, опис програми тренінгу із зазначенням змісту тем, обсягу та тривалості. <i>Рекомендовані курси, тренінги, стажування:</i> https://github.com/DataTalksClub/mlops-zoomcamp https://www.netacad.com/courses/introduction-to-modern-ai?courseLang=en-US Порядок визнання результатів навчання регламентує Положення про визнання результатів навчання у Запорізькому інституті економіки та інформаційних технологій, здобутих шляхом неформальної та/або інформальної освіти (https://www.zieit.edu.ua/pdf/litsenzuvannia-ta-atestatsiia/2023stp68.pdf).
9. Політика освітнього компонента.	Курс передбачає роботу в колективі. Середовище в аудиторії є дружнім, творчим, відкритим до конструктивної критики. Відвідування занять є обов'язковим компонентом оцінювання. За об'єктивних причин (наприклад, хвороба, міжнародне стажування та інше) навчання може відбуватись в дистанційному режимі за погодженням із керівником курсу та презентацією виконаних завдань під час консультації викладача. Усі завдання, передбачені програмою, мають бути виконані у встановлений термін. За використання телефонів і комп'ютерних засобів без дозволу викладача, порушення дисципліни здобувач освіти отримує за заняття 0 балів і зобов'язаний відпрацювати таке заняття. Здобувачам вищої освіти після аудиторних занять надається право підвищувати свій рейтинг лише під час складання заліків (іспитів, за графіком екзаменаційної сесії). <i>Складання заліку:</i> Залік складається у випадку недостатньої кількості балів, набраних в процесі модульного контролю, при наявності не менш ніж 30 балів поточного контролю у вигляді співбесіди з викладачем за списком заздалегідь означених питань. Здобувачі вищої освіти мають право скласти залік автоматично, у випадку, якщо впродовж семестру (поточний + модульний контроль) такі здобувачі набрали 60-100 балів. <i>Політика щодо використання ІІІ та академічної доброчесності:</i>

	Студенти можуть використовувати ШІ як інструмент для пошуку ідей, структурування тексту або редагування граматики. Також, пропонується використовувати штучний інтелект для самостійного вивчення матеріалу, як інструмент роз'яснення складних питань. Надані ШІ рекомендації або пояснення мають бути перевірені в інших джерелах. Проте заборонено копіювати згенерований ШІ текст напряму у свої роботи. Якщо ви використали ШІ під час підготовки завдання, ви зобов'язані: • Зазначити, яку саме модель ШІ ви використовували (наприклад, ChatGPT 4o). • Описати, як саме інструмент допоміг (наприклад, «для генерації ідей до плану»). • Додати текст вашого запиту (промпту) у додатках до роботи. Відповідальність за точність наведених фактів несе студент. Під час роботи над завданнями не допустимо порушення академічної доброчесності: при використанні Інтернет ресурсів та інших джерел інформації студент повинен вказати джерело, використане під час виконання завдання. У разі виявлення факту плагіату здобувач освіти отримує за завдання 0 балів і повинен повторно виконати завдання, які передбачені у силабусі. Списування під час контрольних робіт та екзаменів заборонені (в т.ч. із використанням мобільних пристроїв). Мобільні пристрої дозволяється використовувати лише під час он-лайн тестування. Політика та принципи академічної доброчесності визначені у ПОЛОЖЕННІ про систему забезпечення академічної доброчесності у Запорізькому інституті економіки та інформаційних технологій
--	---

10. Розподіл балів, які отримують здобувачі освіти (відповідає табл. 2 з урахуванням завдань сам.роботи)

Семестр	Модуль	Назва теми	Кількість балів
1	1	Тести за змістовними модулями	
		Змістовний модуль 1	5
		Змістовний модуль 2	5
		Змістовний модуль 3	5
		Змістовний модуль 4	5
		Практична робота №1	10
		Практична робота №2	10
		Практична робота №3	10
		Практична робота №4	10
		Модульні тести	40
	Разом		100

Шкала оцінювання: національна та ЄКТС

Сума балів за всі види навчальної діяльності	Оцінка ECTS	Оцінка за національною шкалою	
		екзамен, курсовий проект (роботи), практика	залік
90 – 100	A	відмінно	зараховано
82-89	B	добре	
74-81	C		
64-73	D	задовільно	
60-63	E		
35-59	FX	незадовільно з можливістю повторного складання	не зараховано з можливістю повторного складання
0-34	F	незадовільно з обов’язковим повторним вивченням дисципліни	не зараховано з обов’язковим повторним вивченням дисципліни

11. Питання до екзамену (заліку):

1. Порівняйте парадигму традиційного програмування (Класичний софт) та машинного навчання з погляду управління даними та визначення логіки.
2. Поясніть сутність процесу оптимізації функцій за допомогою градієнтного спуску в контексті навчання моделей.
3. У чому полягають архітектурні відмінності між біологічним нейроном та математичною моделлю нейрона Маккалокі-Піттса?
4. Опишіть принцип роботи класичного перцептрона та поясніть обмеження його здатності до лінійної класифікації.
5. Розкрийте суть поняття «квадрат компромісів» (The AI Trade-off Square) при проєктуванні інфраструктури для штучного інтелекту.
6. Яким чином вибір архітектури глибокої нейронної мережі впливає на «пляшкове горло» пам'яті (memory bottleneck) та енергоспоживання?
7. Які ключові інженерні виклики постають під час розгортання та виконання моделей глибокого навчання (Deep Learning) у продакшен-середовищі?
8. Порівняйте архітектурні особливості та сферу застосування патернів Transformers та Graph Neural Networks (GNN).
9. Яку роль відіграє метод квантування (Quantization) у досягненні балансу в квадраті компромісів для моделі?
10. Порівняйте структурний та неструктурний прунінг (Pruning) з погляду ефективності використання апаратних ресурсів.
11. Чому традиційні компілятори (наприклад, для C/C++) не підходять для ефективної оптимізації виконання тензорних обчислень III?
12. Що таке операторне злиття (Operator Fusion) у сучасних графових компіляторах і яку проблему воно вирішує?
13. Які технології (наприклад, WebGPU, ONNX Runtime Web) забезпечують виконання інференсу III безпосередньо у веббраузері та які їхні обмеження?
14. Опишіть призначення нативних мобільних фреймворків CoreML (iOS) та NNAPI (Android) у контексті Edge-обчислень.
15. З якими суворими апаратними обмеженнями стикаються інженери під час розгортання систем TinyML на мікроконтролерах?
16. Як саме статичне виділення тензорів оптимізує роботу оперативної пам'яті у TensorFlow Lite for Microcontrollers (TFLM)?
17. Поясніть концепцію MLOps та охарактеризуйте різницю між керуванням версіями коду, даних (за допомогою DVC) і моделей (MLflow/ClearML).

18. У чому полягає відмінність між явищами Data Drift та Concept Drift, і чому їх моніторинг критичний для життєвого циклу моделі?
19. Які переваги для промислового сервінгу надають спеціалізовані сервери виклику моделей (Triton Inference Server, TorchServe) порівняно зі звичайними вебсерверами?
20. Що таке патерн RAG (Retrieval-Augmented Generation) та яка роль векторних баз даних в інженерії сучасних мовних моделей (LLMOps)?

12. Навчально-методичне та інформаційне забезпечення

12.1. Основна:

1. **PyTorch** [Електронний ресурс]. — Режим доступу: <https://pytorch.org>
2. **ONNX (Open Neural Network Exchange)** [Електронний ресурс]. — Режим доступу: <https://onnx.ai>
3. **Triton Inference Server** [Електронний ресурс] / NVIDIA. — Режим доступу: <https://github.com/triton-inference-server/server>
4. **MLOps Zoomcamp** [Електронний ресурс] / DataTalks.Club. — Режим доступу: <https://github.com/DataTalksClub/mlops-zoomcamp>

12.2 Додаткова:

1. Глибовець В. М., Кононець О. А. Застосування графових нейронних мереж для аналізу складних мережових структур. *Проблеми програмування*. 2021. № 2. С. 45–54.
2. Сидор А. Р., Войтович О. П. Оптимізація та розгортання глибоких нейронних мереж на обмежених ресурсах за допомогою інструментарію Intel OpenVINO. *Сучасні комп'ютерні системи та мережі*. 2022. № 1. С. 89–97.
3. Ткаченко Р. О., Дорошенко А. Ю. Низькорівнева оптимізація обчислювальних графів штучних нейронних мереж для графічних процесорів. *Системні дослідження та інформаційні технології*. 2023. № 3. С. 112–121.
4. Козак О. В., Романюк В. О. Особливості застосування технологій TinyML у розподілених системах Інтернету речей. *Вісник Хмельницького національного університету. Серія: Технічні науки*. 2021. № 4. С. 115–122.
5. Запорожець О. В., Шевченко С. М. Федеративне навчання як метод забезпечення конфіденційності даних у розподілених інформаційних системах. *Комп'ютерно-інтегровані технології: освіта, наука, виробництво*. 2022. Вип. 48. С. 67–74.
6. Ткаченко А. М. Використання технології WebGPU для прискорення обчислень у веб-середовищі. *Сучасні інформаційні технології та комп'ютерна інженерія*. 2023. № 2. С. 41–48.