

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

ЗАТВЕРДЖЕНО

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

ПОГОДЖЕНО

Проректор з навчально-методичної роботи

Каріна НЕМАШКАЛО



ЕТИКА ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ

робоча програма навчальної дисципліни (РПНД)

Галузь знань **F Інформаційні технології**
Спеціальність **F6 «Інформаційні системи і технології»**
Освітній рівень **перший (бакалаврський)**

Освітня програма **"Штучний інтелект"**

Статус дисципліни **обов'язкова**
Мова викладання, навчання та оцінювання **українська**

Розробник:
к.е.н., доц.

Ольга ВІЛЬХІВСЬКА

Завідувач кафедри
інформатики та комп'ютерної
техніки

Сергій УДОВЕНКО

Гарант програми

Сергій УДОВЕНКО

Харків
2025

ВСТУП

Сучасний розвиток інформаційних систем і технологій, зокрема сфери штучного інтелекту, нерозривно пов'язаний із впровадженням потужних алгоритмів, які стають основою автоматизації процесів, прийняття рішень, аналізу даних, персоналізації сервісів та інновацій у різних галузях. Штучний інтелект (ШІ) відкриває нові можливості для підвищення ефективності, точності прогнозування, оптимізації ресурсів та створення інтелектуальних систем, але водночас породжує серйозні етичні виклики, пов'язані з упередженістю алгоритмів, порушенням приватності, маніпуляцією поведінкою, дискримінацією, втратою робочих місць, відсутністю прозорості («чорна скринька»), відповідальністю за шкоду та потенційним загрозами фундаментальним правам людини. Етичне використання ШІ вимагає балансу між інноваціями та захистом людської гідності, справедливості й безпеки в цифровому суспільстві.

Навчальна дисципліна "Етика використання штучного інтелекту" є обов'язковою для здобувачів першого (бакалаврського) рівня вищої освіти спеціальності Ф6 "Інформаційні системи і технології", освітня програма "Штучний інтелект". Її зміст спрямований на формування у здобувачів етичної свідомості, критичного мислення та відповідального підходу до розробки, впровадження й експлуатації систем штучного інтелекту.

Метою дисципліни є формування компетентностей щодо етичних принципів, ризиків та регуляторних вимог використання штучного інтелекту, забезпечення дотримання прав людини, прозорості, справедливості та відповідальності в процесі створення та застосування ШІ-систем.

Завданнями навчальної дисципліни є: формування етичних компетентностей для відповідальної розробки та використання штучного інтелекту шляхом ознайомлення з ключовими міжнародними стандартами (Рекомендація UNESCO щодо етики ШІ, EU AI Act, принципи OECD), аналізом основних етичних ризиків (упередженість і дискримінація, порушення приватності даних, відсутність пояснюваності, вплив на зайнятість, маніпуляція та соціальне шкодування), вивченням ризик-орієнтованого підходу (неприйнятний ризик, високий ризик, прозорість для користувачів), порівняльним аналізом реальних кейсів етичних порушень у ШІ (facial recognition, алгоритми кредитного скорингу, чат-боти, generative AI) та практичним опануванням інструментів етичного аудиту й оцінки – від виявлення bias у датасетах і забезпечення fairness до розробки політик прозорості, людського нагляду, відповідальності та підготовки рекомендацій щодо етичного дизайну ШІ-систем для підвищення довіри суспільства й сталого розвитку технологій.

Предметом вивчення виступає етичний вимір розробки, впровадження та застосування штучного інтелекту в інформаційних системах, що включає принципи відповідального ШІ, оцінку ризиків, регуляторні рамки та механізми захисту прав і свобод людини.

Об'єктом вивчення є системи штучного інтелекту та їх вплив на суспільство, економіку й індивідуальні права в цифровому середовищі з урахуванням етичних, правових та соціальних аспектів.

У межах дисципліни здобувачі знайомляться з основами етики ШІ, методами виявлення та мінімізації ризиків (bias detection, privacy by design, explainable AI), міжнародними та європейськими нормативними документами (UNESCO Recommendation, EU AI Act), а також практичними підходами до етичного впровадження ШІ (аудит алгоритмів, оцінка впливу, політики governance). Особлива увага приділяється аналізу етичних кейсів у сфері інформаційних технологій (рекомендаційні системи, автономні рішення, generative models) та формуванню навичок балансування інновацій з етичними нормами. Вивчення дисципліни дозволяє майбутнім фахівцям не лише зрозуміти теоретичні засади етики ШІ, але й застосовувати відповідальні практики для мінімізації шкоди, забезпечення справедливості, прозорості та збереження людського контролю, сприяючи стійкому та довірливому розвитку штучного інтелекту в інформаційному суспільстві.

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

Таблиця 1

Результати навчання та компетентності, які формує навчальна дисципліна

Результати навчання	Компетентності, якими повинен оволодіти здобувач вищої освіти
ПР4	ІК, К31, К33, К39, КС2, КС5, КС6
ПР10	ІК, К310
ПР11	ІК, К35, К311
ПР13	ІК, КС15, КС16, КС19

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

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

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

ПР 13. Аналізувати та виявляти ризики неетичного використання інформаційних моделей штучного інтелекту в контексті вирішення практичної задачі, пропонувати методи уникання цих ризиків.

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

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

КЗ 3. Здатність до розуміння предметної області та професійної діяльності.

КЗ 5. Здатність вчитися і оволодівати сучасними знаннями.

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

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

КЗ 11. Здатність ухвалювати рішення та діяти, дотримуючись принципу неприпустимості корупції та будь-яких проявів недоброчесності.

КС 2. Здатність застосовувати стандарти в області інформаційних систем та технологій при розробці функціональних профілів, побудові та інтеграції систем, продуктів, сервісів і елементів інфраструктури організації.

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

КС 6. Здатність використовувати сучасні інформаційні системи та технології (виробничі, підтримки прийняття рішень, інтелектуального аналізу даних та інші), методики й техніки кібербезпеки під час виконання функціональних завдань та обов'язків.

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

КС 16. Здатність формувати набори даних для навчання інформаційних моделей штучного інтелекту, проектувати нові моделі або обирати відповідні з тих, що вже існують

КС 19. Здатність оцінювати ризики неетичного використання інформаційних моделей штучного інтелекту для вирішення прикладних задач.

ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Зміст навчальної дисципліни

Змістовий модуль 1 Основи етики штучного інтелекту: поняття, виклики та відповідальність.

Тема 1. Визначення етики ШІ та історія терміну.

1.1. Визначення етики штучного інтелекту та її значення в сучасному світі.

Поняття етики штучного інтелекту (AI ethics) як галузі, що вивчає моральні принципи, ризики та відповідальність у розробці, впровадженні та використанні систем ШІ. Значення етики ШІ для забезпечення справедливості, прозорості, захисту прав людини, запобігання упередженості, приватності даних та

уникнення шкоди суспільству. Вплив етичних аспектів на довіру до технологій, регуляторні рамки та стале використання ШІ в інформаційних системах, економіці, медицині, освіті та інших сферах.

1.2. Історія розвитку етики ШІ: ключові етапи та віхи.

Основні етапи еволюції етичних роздумів про інтелектуальні машини: від ранніх філософських ідей (Норберт Вінер, 1948-1950, попередження про моральні наслідки автономних систем) та класичних науково-фантастичних рамок (Три закони робототехніки Айзека Азімова, 1942) до народження сучасної етики ШІ. Ключові події: Dartmouth Conference (1956, початок терміну "штучний інтелект" та перші дискусії про вплив), ELIZA (1966, етичні питання антропоморфізму), Asilomar Conference on Beneficial AI (2017, фокус на безпечному та корисному ШІ), Рекомендація UNESCO щодо етики ШІ (2021, перший глобальний стандарт), EU AI Act (2024-2026, ризик-орієнтоване регулювання). Перехід від теоретичних дискусій до практичних викликів (bias, explainability, job displacement, existential risks) у 2010-2020-х роках.

1.3. Приклади ключових документів, ініціатив та сучасних підходів до етики ШІ.

Огляд провідних рамок та принципів: Asilomar AI Principles (2017), OECD AI Principles (2019), UNESCO Recommendation on the Ethics of AI (2021), EU AI Act (high-risk classification, transparency requirements), IEEE Ethically Aligned Design. Порівняння підходів: від загальних принципів (fairness, accountability, transparency) до конкретних механізмів (bias auditing, human oversight, value alignment). Вибір та адаптація етичних рамок залежно від типу ШІ-систем (низький/високий ризик, генеративні моделі, автономні рішення) та контексту застосування в інформаційних технологіях. Аналіз реальних кейсів етичних порушень та успішних практик відповідального ШІ для ілюстрації еволюції терміну та його практичного значення.

Тема 2. Етичні виклики ШІ.

2.1. Основні етичні виклики та ризики застосування штучного інтелекту

Поняття етичних викликів ШІ як потенційних негативних наслідків для окремої людини, суспільства та довкілля. Ключові проблеми: упередженість алгоритмів (bias) та дискримінація, порушення приватності персональних даних, відсутність прозорості та пояснюваності рішень ("чорна скринька"), маніпуляція поведінкою людини, поширення дезінформації (deepfakes, галюцинації моделей), автоматизоване заміщення робочих місць та зростання соціальної нерівності, екологічний вплив (високе енергоспоживання та вуглецевий слід), ризики автономних систем (agentic AI) та потенційні загрози безпеці й існуванню людства. Вплив цих викликів на рівень довіри до технологій та необхідність пошуку балансу між інноваціями та захистом фундаментальних прав людини.

2.2. Сучасні тенденції етичних ризиків ШІ

Аналіз актуальних трендів: зростання фокусу на захисті від автономних агентів (agentic guardrails), концентрація влади в руках кількох великих корпорацій, посилення загроз від генеративного ШІ (масштабна дезінформація, маніпуляція ідентичністю), фрагментація глобального регулювання та відставання правових норм від темпів розвитку технологій, ризики для

когнітивного розвитку людини (особливо в освіті), екологічні та соціальні наслідки масштабного використання ШІ, перехід від теоретичних дискусій до реальних судових справ та регуляторних санкцій (аудити упередженості, відповідальність за автономні помилки). Приклади впливу на суспільство: упередженість у системах підбору персоналу, масове стеження, автономна зброя.

2.3. Приклади реальних етичних порушень та кейсів

Огляд відомих випадків: упередженість у системах розпізнавання обличчя та кредитного скорингу, маніпуляція емоціями через emotional AI, масове поширення deepfakes та фейкової інформації, величезний екологічний слід великих мовних моделей, заміщення робочих місць у кол-центрах та програмуванні, порушення приватності в рекомендаційних системах. Порівняння наслідків та реакцій (штрафи регуляторів, суспільний осуд, добровільне відкликання продуктів). Аналіз уроків для майбутньої розробки ШІ-систем та необхідності проактивного етичного дизайну.

Тема 3. Відповідальність у використанні ШІ.

3.1. Поняття відповідальності та підзвітності в ШІ.

Визначення відповідальності (responsibility, accountability, liability) у контексті ШІ: хто несе відповідальність за шкоду – розробник, власник, користувач чи сама автономна система? Відмінність між моральною, юридичною та технічною відповідальністю. Значення людського нагляду (human oversight), пояснюваності та механізмів відшкодування для забезпечення підзвітності.

3.2. Сучасні рамки та підходи до відповідальності.

Огляд ключових концепцій: ризик-орієнтовані підходи (EU AI Act), обов'язкові механізми підзвітності (AI Accountability Framework), відповідальність за автономних агентів, індемніфікація в контрактах, аудити та системи управління. Принципи: запобігання, прозорість, відшкодування, постійний людський контроль. Приклади розподілу відповідальності в автономних системах.

3.3. Практичні механізми забезпечення відповідальності.

Аналіз інструментів: оцінка ризиків, звітування про інциденти, аудити упередженості, документація життєвого циклу, політики прийнятого використання. Кейси: відповідальність за галюцинації, автономні агенти, рішення щодо працевлаштування. Рекомендації для організацій щодо впровадження підзвітності на етапі дизайну (accountability-by-design).

Тема 4. Етика роботів.

4.1. Поняття roboethics та її відмінності від загальної етики ШІ.

Визначення roboethics як окремої підгалузі, що зосереджується на етичних аспектах фізичних роботів (промислових, сервісних, медичних, автономних). Зв'язок із класичними ідеями та сучасними викликами (взаємодія людина-робот, заміщення людської праці).

4.2. Ключові етичні питання етики роботів.

Основні проблеми: автономія та моральне прийняття рішень роботами, вплив на зайнятість та соціальні структури, етика догляду (роботи-компаньйони, догляд за літніми), хірургічні роботи та відповідальність за помилки, дизайн

етики (узгодження цінностей), потенційні права роботів (у разі досягнення свідомості), етика взаємодії (обслуговування клієнтів роботами). Вплив на суспільство та необхідність етичного дизайну.

4.3. Приклади та кейси етики роботів.

Огляд реальних застосувань: промислові роботи та благополуччя працівників, сервісні роботи в HoReCa та медицині, військові роботи, роботи-компаньйони та сексуальні роботи. Аналіз етичних дилем та регуляторних підходів (людський нагляд, стандарти безпеки).

Тема 5. Популярні принципи етичного ШІ.

5.1. Основні принципи етичного ШІ за міжнародними стандартами.

Огляд ключових принципів: Рекомендація UNESCO (права людини та гідність, прозорість, справедливість, сталість), принципи OECD (інклюзивне зростання, надійність, підзвітність), EU AI Act (ризико-орієнтований підхід, пропорційність).

5.2. Порівняння глобальних етичних рамок.

Аналіз спільного та відмінного: справедливість, приватність, пояснюваність, людиноцентричний ШІ. Вплив на національні політики.

5.3. Практичне застосування принципів.

Приклади впровадження: етичні аудити, узгодження цінностей, системи управління.

Тема 6. Основні етичні загрози використання ШІ.

6.1. Класифікація етичних загроз.

Категорії: упередженість та дискримінація, ерозія приватності, дезінформація, заміщення робочих місць, екологічний вплив, загрози безпеці, екзистенційні ризики.

6.2. Сучасні загрози.

Фокус на автономних агентах, генеративних моделях, концентрації влади, питаннях суверенітету.

6.3. Стратегії мінімізації загроз.

Запобігання: виявлення упередженості, privacy by design, відповідність регуляціям.

Тема 7. Етика використання ШІ в освіті.

7.1. Можливості та ризики ШІ в освіті.

Персоналізоване навчання проти загроз когнітивному розвитку, плагіат, нерівний доступ.

7.2. Етичні виклики та рекомендації. Практичні підходи.

Академічна доброчесність, автономія викладача, рекомендації UNESCO, наслідки EU AI Act. AI-грамотність, етична інтеграція, людиноцентричні практики.

Тема 8. Рекомендації для бізнесу та суспільства.

8.1. Рекомендації для бізнесу, суспільства та політики. Майбутні перспективи.

Фреймворки управління ШІ, політики, аудити, етичні комітети. Глобальна гармонізація, освіта з етики ШІ, інклюзивне управління.

Вбудовування етики як інфраструктури, перехід до обов'язкових стандартів.

Перелік лабораторних занять за навчальною дисципліною наведено в табл. 2

Таблиця 2

Перелік лабораторних занять

Назва теми та / або завдання	Зміст
Тема 1. Завдання 1. Виявлення та аналіз упередженості (bias) в датасетах та моделях	Практичне виявлення дискримінації в даних та моделях за допомогою метрик справедливості та пошук методів для їхнього корегування
Тема 2. Завдання 2. Оцінка пояснюваності моделей (Explainable AI)	Дослідження «чорних скриньок» за допомогою алгоритмів інтерпретації (SHAP, LIME) для розуміння логіки прийняття рішень моделлю
Тема 3. Завдання 3. Аналіз приватності даних та атаки на конфіденційність	Тестування стійкості моделей до витоку персональних даних та впровадження методів захисту, таких як диференціальна приватність
Тема 4. Завдання 4. Етичний аудит генеративної моделі	Критична оцінка ризиків генеративного ШІ через тестування на стереотипи, шкідливий контент та розробка запобіжних фільтрів
Тема 5, 6. Завдання 5. Аналіз етичних кейсів та розробка політики використання ШІ	Ретроспективний аналіз відомих помилок ШІ та створення регуляторної політики організації згідно з міжнародними стандартами
Тема 7, 8. Завдання 6. Сценарійний аналіз: етичне рішення в конкретному бізнес-кейсі	Оцінка ризиків впровадження ШІ-систем у конкретних бізнес-ситуаціях та прийняття обґрунтованого рішення про їхню доцільність

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

Таблиця 3

Перелік самостійної роботи

Назва теми та / або завдання	Зміст
Тема 1-8	Вивчення лекційного матеріалу
Тема 1-8	Підготовка до лабораторних робіт
Тема 1-8	Виконання індивідуальних завдань

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

МЕТОДИ НАВЧАННЯ

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

- словесні (лекції за темами 1-8), елементи проблемних лекції (теми 1-8);
- наочні (демонстрації (теми 1-8);
- практичні (лабораторні заняття за темами 1-8).

В умовах змішаної форми навчання подання лекційного матеріалу та/або проведення лабораторних занять та групових та індивідуальних консультацій відбувається з використанням платформи Zoom, в умовах звичайної аудиторної форми заняття проводяться очно, в аудиторіях та комп'ютерних залах.

ФОРМИ ТА МЕТОДИ ОЦІНЮВАННЯ

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

Поточний контроль здійснюється під час проведення лекційних та лабораторних занять і має на меті перевірку рівня підготовленості здобувача вищої освіти до виконання конкретної роботи і оцінюється сумою набраних балів:

– для дисциплін з формою семестрового контролю залік: максимальна сума – 100 балів; мінімальна сума – 60 балів.

Підсумковий контроль включає семестровий контроль та атестацію здобувача вищої освіти.

Семестровий контроль проводиться у формах диференційованого заліку або заліку.

Підсумкова оцінка за навчальною дисципліною визначається:

– для дисциплін з формою семестрового контролю залік – сумуванням всіх балів, отриманих під час поточного контролю.

Під час викладання навчальної дисципліни використовуються наступні контрольні заходи:

Поточний контроль: захист лабораторних робіт (60 балів), письмові контрольні роботи (20 балів), тести (20 балів).

Семестровий контроль: Залік

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

РЕКОМЕНДОВАНА ЛІТЕРАТУРА

Основна

1. Когут Ю. І. Етика штучного інтелекту. – Київ : Кондор, 2021. 540 с.
2. Шевченко А. С. Правове регулювання штучного інтелекту: світовий досвід та перспективи для України. – Київ : Юрінком Інтер, 2023. – 240 с.

Додаткова

3. Molnar C. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. 2nd ed. Leanpub, 2022. 454 p.

4. Barocas S., Hardt M., Narayanan A. Fairness and Machine Learning: Limitations and Opportunities. MIT Press, 2023. 272 p.
5. Dwork C., Roth A. The Algorithmic Foundations of Differential Privacy. Foundations and Trends in Theoretical Computer Science. 2014. Vol. 9, no. 3–4. P. 211–407.
6. Vilkhivska O. Ethical aspects of using Artificial Intelligence : challenges and prospects Академічні візії. – 2025. – Вип.40. – С. 1-5 – [Електронний ресурс]. – Режим доступу :<https://repository.hneu.edu.ua/handle/123456789/36256>.

Інформаційні ресурси

7. Сайт персональних навчальних систем ХНЕУ ім. С. Кузнеця [Електронний ресурс]. – Режим доступу : <https://pns.hneu.edu.ua/mod/folder/view.php?id=710665>.