

Міністерство освіти і науки України
Національний аерокосмічний університет ім. М. Є. Жуковського
«Харківський авіаційний інститут»



ЗАТВЕРДЖУЮ

Проректор з наукової роботи

В. В. Павліков

(ініціали та прізвище)

08 2020 р.

Відділ аспірантури і докторантури

**РОБОЧА ПРОГРАМА ВИБІРКОВОЇ
НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**

Теорія і методи data science і штучного інтелекту

(назва навчальної дисципліни)

Галузь знань: 12 "Інформаційні технології"

(шифр і найменування галузі знань)

Спеціальність: 123 "Комп'ютерна інженерія"

(код та найменування спеціальності)

Освітньо-наукова програма "Комп'ютерна інженерія"

(найменування)

Рівень вищої освіти: третій (освітньо-науковий)

Форма навчання: денна

Харків 2020 рік

**РОБОЧА ПРОГРАМА
ВИБІРКОВОЇ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**

Теорія і методи data science і штучного інтелекту
(назва дисципліни)

для здобувачів за спеціальністю 123 "Комп'ютерна інженерія"
освітньо-наукової програми Комп'ютерна інженерія

«26» серпня 2020 р., – 11 с.

Розробник: доцент, к.т.н., доцент
(посада, науковий ступінь та вчене звання)


(підпис)

Фесенко Г. В.
(прізвище та ініціали)

Гарант ОНП доцент, к.т.н.
(посада, науковий ступінь та вчене звання)


(підпис)

Бабешко С.В.
(прізвище та ініціали)

Протокол №1 від «27» серпня 2020 р. засідання кафедри № 503

Завідувач кафедри д.т.н., професор
(науковий ступінь та вчене звання)


(підпис)

Харченко В. С.
(прізвище та ініціали)

ПОГОДЖЕНО:

Завідувач відділу
аспірантури і докторантури



В. Б. Селевко

Голова наукового товариства
студентів, аспірантів,
докторантів і молодих вчених



Т. П. Старовойт

1. Опис навчальної дисципліни

Найменування показників	Галузь знань, спеціальність, освітня програма, рівень вищої освіти	Характеристика навчальної дисципліни (денна форма навчання)
Кількість кредитів – 5	Галузь знань <u>12 "Інформаційні технології"</u> (шифр та найменування) Спеціальність <u>123 "Комп'ютерна інженерія"</u> (код та найменування) Освітньо-наукова програма <u>"Комп'ютерна інженерія"</u> (найменування) Рівень вищої освіти: <u>третій (освітньо-науковий)</u>	Вибіркова
Кількість модулів – 1		Навчальний рік
Кількість змістовних модулів – 2		2020/ 2021
<u>Індивідуальне завдання</u> - немає (назва)		Семестр
Загальна кількість годин: 68/150		<u>4-й</u>
		Лекції *
Кількість тижневих годин для денної форми навчання: аудиторних – 4, самостійної роботи аспіранта – 4,8		34 годин
		Практичні, семінарські*
		0 годин
		Лабораторні*
	34 години	
	Самостійна робота	
	82 години	
	Вид контролю	
	Іспит	

Співвідношення кількості годин аудиторних занять до самостійної роботи становить: **68/82.**

*Аудиторне навантаження може бути зменшене або збільшене на одну годину в залежності від розкладу занять.

2. Мета та завдання навчальної дисципліни

1. Мета вивчення: формування знань, вмінь та навичок, необхідних для застосовування в наукових дослідженнях: сучасних інструментів і технологій пошуку, оброблення та аналізу даних великого обсягу та/або складної структури; систем штучного інтелекту сумісно із сучасними методами і засобів програмування.

2. Завдання: комплексне застосування технологій та методів обробки, зберігання та аналізу даних великого обсягу та/або складної структури, а також методів та моделей штучного інтелекту.

3. Програмні компетентності. Дисципліна має допомогти сформувати у аспірантів такі компетентності:

- здатність до абстрактного мислення, аналізу та синтезу;
- здатність до пошуку, оброблення та аналізу інформації з різних джерел;
- здатність застосовувати сучасні інформаційні технології, бази даних та інші електронні ресурси, спеціалізоване програмне забезпечення у науковій та навчальній діяльності;
- здатність до продукування нових ідей і розв'язання комплексних проблем у галузі інформаційних технологій, а також до застосування сучасних методологій, методів та інструментів педагогічної та наукової діяльності в комп'ютерній інженерії.

4. Програмні результати навчання. В результаті вивчення дисципліни аспіранти мають досягти такі програмні результати навчання:

- застосовувати сучасні інструменти і технології пошуку, оброблення та аналізу інформації, зокрема, статистичні методи аналізу даних великого обсягу та/або складної структури, спеціалізовані бази даних та інформаційні системи;
- знати, розуміти та вміти застосовувати методи та засоби створення інформаційних технологій та програмного забезпечення розподілених систем, Інтернету речей, хмарних обчислень, систем штучного інтелекту, віртуальної реальності у предметних областях різних галузей, в тому числі в аерокосмічній галузі.

5. Міждисциплінарні зв'язки. Дисципліна є вибірковою компонентою освітньо-наукової програми «Комп'ютерна інженерія» і базується на знаннях, отриманих під час вивчення дисципліни «ІТ в практиці наукових досліджень», «Теорія і методи зеленої ІТ-інженерії», «Наукові англомовні комунікації», «Теорія планування експерименту», що є обов'язковими компонентами.

Матеріал, засвоєний під час вивчення цієї дисципліни, а також є базою для підготовки дисертаційної роботи.

3. Програма навчальної дисципліни

Модуль 1

Змістовний модуль 1. Теорія і методи data science

Тема 1. Базові поняття великих даних.

Набори та аналіз даних. Дескриптивна, діагностична, прогностична та прескриптивна аналітика з використанням великих даних. Бізнес аналітика. Ключові показники ефективності. Характеристики великих даних (об'єм, швидкість, різноманіття, достовірність, цінність). Різні типи даних (структуровані, неструктуровані, слабо структуровані та метадані).

Тема 2. Питання планування та бізнес-аналітика в наукових дослідженнях

Організаційні передумови. Набуття даних. Конфіденційність. Безпека. Походження. Обмеження підтримки у реальному часі. Особливі проблеми продуктивності. Особливі вимоги до керівництва. Методологія. Великі дані і хмарні технології. Життєвий цикл великих даних. Обробка транзакцій в режимі реального часу (OLTP). Аналітична обробка у реальному часі (OLAP). Вилучення, перетворення та завантаження (ETL). Сховища даних. Вітрини даних. Традиційна бізнес-аналітика та бізнес-аналітика для великих даних.

Тема 3. Концепції зберігання та обробки великих даних в наукових дослідженнях.

Кластери. Файлові системи та розподілені файлові системи. NoSQL. Шардінг. Реплікація. Шардінг і реплікація. Теорема CAP. ACID. BASE. Паралельна обробка даних. Розподілена обробка даних. Hadoop. Обробка робочих завдань (пакетна обробка). Транзакційна обробка. Кластер. Обробка у пакетному режимі. Обробка в режимі реального часу (об'єм, узгодженість, швидкість (ОУШ), обробка потоку подій, обробка складних подій в режимі реального часу і ОУШ), обробка великих даних в режимі реального часу і MapReduce. Apache Spark.

Тема 4. Технології зберігання великих даних та основні методи їх аналізу в наукових дослідженнях.

Дискові пристрої зберігання, бази даних NoSQL. Системи зберігання у оперативній пам'яті (In-memory Data Grid, наскрізне читання, наскрізний запис, відкладений запис, оновлення, In-memory Databases). Кількісний аналіз. Якісний аналіз. Data Mining. Статистичний аналіз (A/B тестування, кореляція, регресія). Машинне навчання (класифікація, кластеризація, виявлення викидів, фільтрація). Семантичний аналіз (обробка природної мови, обробка тексту,

аналіз емоціонального забарвлення висловлювань). Візуальний аналіз (кольорові карти, часові ряди, мережеві графи). Співвідношення просторових даних.

Модульний контроль.

Змістовний модуль 2. Теорія і методи штучного інтелекту.

Тема 5. Базові поняття штучного інтелекту та інтелектуальні системи.

Визначення та історія виникнення. Приклади інтелектуальних задач. Тест Тьюринга. Керування складними системами: алгоритмічний та декларативний підходи до керування; формалізація понять алгоритмічності та декларативності. Квазіалгоритми. Характеристика інтелектуальних систем з точки зору кібернетики: означення інтелектуальної системи; типова схема функціонування інтелектуальної системи.

Тема 6. Подання знань в інтелектуальних системах під час наукових досліджень

Підходи до подання знань. Вербально-дедуктивне визначення знань. Експертні системи. Дані та знання. Зв'язки між інформаційними одиницями. Проблема винятків. Властивості та моделі знань. Неоднорідність знань. Області і рівні знань. База знань як об'єднання простіших одиниць. Бінарні предикати і тріада "об'єкт—атрибут—значення".

Тема 7. Мережеві та фреймові моделі знань, продукційні моделі та їх застосування в наукових дослідженнях.

Семантичні мережі. Фрейми. Зв'язок між семантичними мережами та фреймами. Загальна характеристика продукційних моделей. Продукції та мережі виведення. Пряме та зворотне виведення. Типові дисципліни виконання продукцій. Основні стратегії вирішення конфліктів у продукційних системах.

Тема 8. Конекціоністські моделі й методи в наукових дослідженнях.

Загальна характеристика конекціоністського підходу та його місце в теорії інтелектуальних систем. Модель штучного нейрона. Архітектура штучних нейронних мереж: штучні нейронні мережі прямого та зворотного поширення, повнозв'язані штучні нейронні мережі. Навчання штучних нейронних мереж: правило навчання Гібба, дельта-правило, градієнтні методи навчання. Одношаровий перцептрон: будова та навчання.

Модульний контроль.

4. Структура навчальної дисципліни

Назва змістовного модуля і тем	Кількість годин				
	Усього	У тому числі			
		л	п	лаб.	с. р.
1	2	3	4	5	6
Модуль 1					
Змістовний модуль 1. Теорія і методи data science.					
Тема 1. Базові поняття великих даних.	18	4		4	10
Тема 2. Питання планування та бізнес-аналітика в наукових дослідженнях	18	4		4	10
Тема 3. Концепції зберігання та обробки великих даних в наукових дослідженнях..	18	4		4	10
Тема 4. Технології зберігання великих даних та основні методи їх аналізу в наукових дослідженнях.	20	4		5	11
Модульний контроль.	1			1	
Разом за змістовним модулем 1	75	16		18	41
Змістовний модуль 2. Теорія і методи штучного інтелекту.					
Тема 5. Базові поняття штучного інтелекту та інтелектуальні системи.	18	4		4	10
Тема 6. Подання знань в інтелектуальних системах під час наукових досліджень.	18	4		4	10
Тема 7. Мережеві та фреймові моделі знань, продукційні моделі та їх застосування в наукових дослідженнях.	18	4		4	10
Тема 8. Конекціоністські моделі й методи в наукових дослідженнях.	20	6		3	11
Модульний контроль	1			1	
Разом за змістовним модулем 2	75	18		16	41
Усього годин за дисципліною	150	34		34	82

5. Теми семінарських занять

№ п/п	Назва теми	Кількість годин
1	<i>Не передбачено</i>	
2		
	Разом	

6. Теми практичних занять

№ п/п	Назва теми	Кількість годин
1	<i>Не передбачено</i>	
2		
	Разом	

7. Теми лабораторних занять

№ п/п	Назва теми	Кількість годин
1	Наукові дослідження з використанням Spark: програмування операцій зі стійкими розподіленими наборами даних (Resilient Distributed Datasets, RDD)	4
2	Наукові дослідження з використанням Spark: створення та робота з DataFrames	4
3	Наукові дослідження з використанням Spark GraphX: робота з графами	4
4	Наукові дослідження з використанням Spark MLlib і пакета spark.ml: реалізація алгоритмів машинного навчання для статистичних обчислень і перевірки гіпотез	6
5	Наукові дослідження з використанням модуля Spark SQL: робота із вбудованими функціями	4
6	Наукові дослідження з використанням Spark MLlib: Transformers and Estimators	4
7	Наукові дослідження з використанням Spark MLlib: Supervised Learning (навчання з учителем)	4
8	Наукові дослідження з використанням Spark MLlib: Recommendation Engines	4
	Разом	34

8. Самостійна робота

№ п/п	Назва теми	Кількість годин
1	Загальні принципи аналізу даних за допомогою Spark	10
2	Особливості роботи зі стійкими розподіленими наборами даних (RDD)	10
3	Особливості роботи с парами «ключ-значення»	10
4	Пакет для роботи зі структурованими даними Spark SQL	11

5	Особливості роботи з великими даним в Spark MLlib: основні відомості про використання машинного навчання для обробки великих даних	10
6	Особливості роботи з великими даним в Spark MLlib: ML Pipelines, Feature Extraction	10
7	Особливості роботи з великими даним в Spark MLlib: Transformation, and Selection, Evaluation Metrics	10
8	Особливості роботи з великими даним в Spark MLlib: навчання з учителем (Supervised Learning)	11
	Разом	82

9. Індивідуальні завдання

Не передбачено навчальним планом

10. Методи навчання

Проведення аудиторних лекцій, практичних занять, консультацій, а також самостійна робота аспірантів за відповідними матеріалами.

11. Методи контролю

Проведення поточного контролю, письмового модульного контролю, підсумковий контроль у вигляді іспиту.

12. Критерії оцінювання та розподіл балів, які отримують аспіранти

12.1. Розподіл балів, які отримують аспіранти (кількісні критерії оцінювання)

Складові навчальної роботи	Бали за одне заняття (завдання)	Кількість занять (завдань)	Сумарна кількість балів
Змістовний модуль 1			
Робота на лекціях	0...3	4	0...12
Виконання лабораторних робіт	0...5	4	0...20
Модульний контроль	0...18	1	0...18
Змістовний модуль 2			
Робота на лекціях	0...3	4	0...12
Виконання лабораторних робіт	0...5	4	0...20
Модульний контроль	0...18	1	0...18
Усього за семестр			0...100

Семестровий контроль у вигляді іспиту проводиться у разі відмови аспіранта від балів поточного тестування й за наявності допуску до іспиту. Під час складання семестрового іспиту аспірант має можливість отримати максимум 100 балів.

Білет для іспиту складається з одного теоретичного та одного практичного запитань, максимальна кількість за кожне із запитань, складає 50 балів.

12.2. Якісні критерії оцінювання

Необхідний обсяг знань для одержання позитивної оцінки:

- знати базові поняття великих даних та штучного інтелекту;
- знати основні концепції зберігання та обробки великих даних;
- знати основні підходи щодо подання знань в інтелектуальних системах;
- знати основні мережеві та фреймові моделі знань.

Необхідний обсяг вмінь для одержання позитивної оцінки:

- уміти здійснювати програмування операцій зі стійкими розподіленими наборами даних;
- уміти працювати з DataFrames та графами з використанням Spark GraphX;
- уміти здійснювати статистичні обчислення і перевірку гіпотез з використанням Spark MLlib і пакета spark.ml.

12.3 Критерії оцінювання роботи аспіранта протягом семестру

Задовільно (60-74). Показати мінімум знань та умінь, виконати не менше 90% завдань лабораторних робіт. Знати базові поняття, що стосуються обробки, зберігання та аналізу великих даних, а також застосування систем штучного інтелекту

Добре (75-89). Твердо знати теоретичний мінімум, опанувати матеріали всіх лекцій і виконати не менше 90% завдань лабораторних робіт. Показати вміння виконувати всі практичні завдання в обумовлений викладачем строк з обґрунтуванням рішень та заходів, які наведено у методичних посібниках з практичних занять.

Відмінно (90-100). Повно знати основний та додатковий матеріал. Знати усі теми. Орієнтуватися у підручниках та посібниках. Досконально знати усі розглянуті концепції і технології обробки, зберігання та аналізу великих даних, а також моделі і методи штучного інтелекту.

Шкала оцінювання: бальна і традиційна

Сума балів	Оцінка за традиційною шкалою	
	Іспит, диференційований залік	Залік
90 – 100	Відмінно	Зараховано
75 – 89	Добре	
60 – 74	Задовільно	
0 – 59	Незадовільно	Не зараховано

13. Методичне забезпечення

1. Методичні матеріали з дисципліни «Теорія і методи data science і штучного інтелекту» [Електронний ресурс]. URL: <https://mentor.khai.edu/course/view.php?id=5307>

14. Рекомендована література

Базова

1. Системи штучного інтелекту в плануванні, моделюванні та управлінні : підруч. для студ. вищ. навч. закл. / Л. С. Ямпольський, Б. П. Ткач, О. І. Лісовиченко. К. : ДП «Вид. дім «Персонал», 2011. 544 с.
2. Artificial Intelligence in Society. Paris : OECD Publishing, 2019. 152 p.
3. Furht B. Big Data technologies and applications / B. Furht, F. Villanustre. Cham : Springer Publishing, 2016. 400 p.
4. Big Data : a primer / Н. Mohanty, P. Bhuyan, D. Chenthati (Eds.). New Delhi : Springer India Publishing, 2016. 400 p.

Допоміжна

1. Ерл Т., Хаттак В., Булер П. Основи Big Data: концепції, алгоритми та технології. Дніпро: Баланс Бізнес Букс, 2018. 320 с.
2. Duvvuri S, Singhal B. Spark for data science. Birmingham – Mumbai : Packt Publ., 2016. 322 p.
3. Zečević P, Bonaći M. Spark-in-action. Shelter Island : Manning Publ., 2017. 443 p.
4. Malak M. S., East R. Spark GraphX in action. Shelter Island : Manning Publ., 2016. 262 p.

15. Інформаційні ресурси

1. What is Artificial Intelligence? How Does AI Work? [Electronic resource]. URL: <https://builtin.com/artificial-intelligence> (date of request 22.08.2019).
2. Association for the Advancement of Artificial Intelligence [Electronic resource]. URL: <https://builtin.com/artificial-intelligence> (date of request 22.08.2019).
3. Курси по великим даним від Coursera [Електронний ресурс]. URL: <https://www.coursera.org/learn/real-time-streaming-big-data?specialization=big-data-engineering> (дата звернення 22.08.2019).