NAAIC

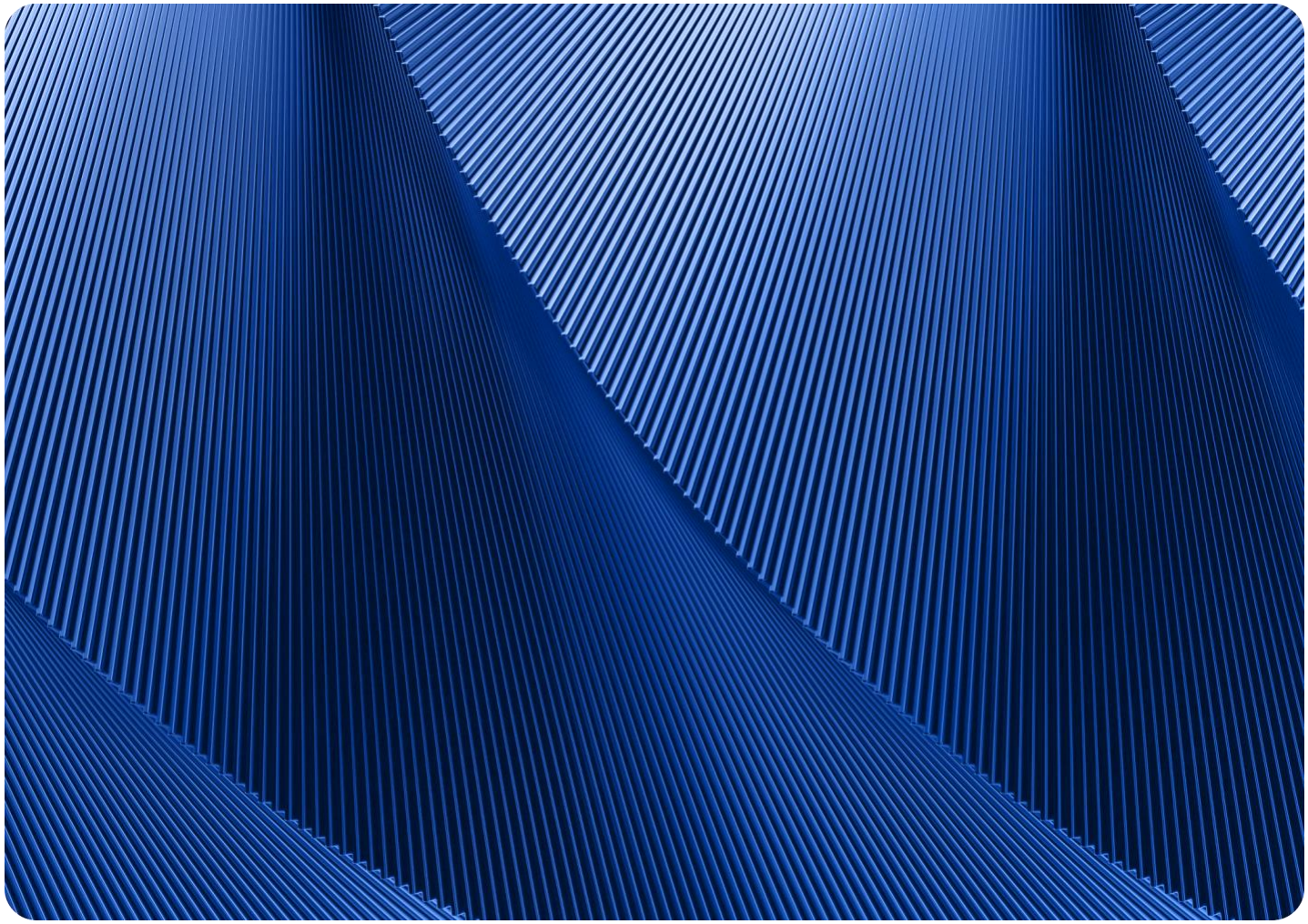
Resource provided by

HCC | HOUSTON COMMUNITY COLLEGE

Course Syllabus

Data Science for AI and Robotics

This is an intermediate course that explores essential data science principles for artificial intelligence and robotics applications.



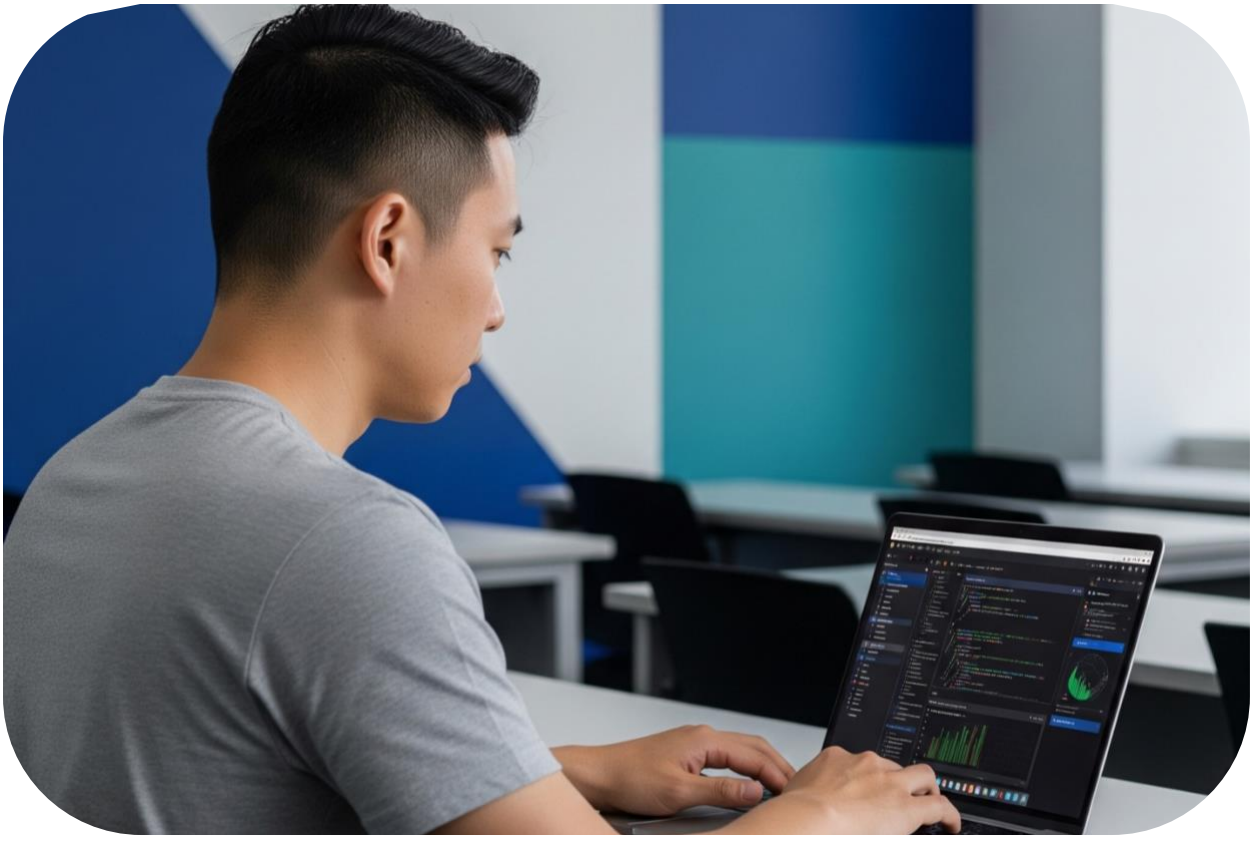
NAAIC

Table of Contents

Course Description

This is an intermediate course that explores essential data science principles for artificial intelligence and robotics applications. The course addresses fundamental challenges in preparing and optimizing information for AI systems across various domains. Students will develop practical skills with contemporary methodologies while building adaptable knowledge that remains relevant as technologies evolve. The course balances technical implementation with ethical considerations, providing a foundation that remains applicable regardless of future advancements in the field.

Prerequisites:
Python programming language
Machine Learning course
Natural Language Processing course



NAAIC

Course Competencies

Competency 1
Understand the evolution of data requirements across different AI paradigms
Competency 2
Master techniques for collecting, preprocessing, and transforming data for various AI applications
Competency 3
Develop practical skills in feature engineering and selection for different model types
Competency 4
Build and optimize data pipelines for AI systems at both experimental and production scales
Competency 5
Implement cloud-based data solutions for enterprise AI applications
Competency 6
Evaluate AI model performance using appropriate metrics and optimization techniques
Competency 7
Generate synthetic data and apply augmentation techniques to address data limitations

NAAIC

Instructional Resources

No prescribed textbook is required due to the rapidly evolving nature of artificial intelligence and related fields. Instructors will curate up-to-date articles, papers, videos, and other resources, available via Canvas, to provide the latest advancements and diverse perspectives. This approach ensures flexibility to tailor content to class needs and incorporate emerging topics.

Platforms and Tools:
Kaggle, GitHub, Google Colab, Amazon SageMaker, VS Code, Hugging Face

Recommended Reading:
See Supplemental Books (Optional) and Supplemental Papers (Optional) sections.

Grading Schema

Assignment Type	Percentage of Grade
General Assignments	20%
Labs/Hands-On Activities	25%
Midterm Project	20%
Capstone Project	25%
Course Portfolio	10%

Supplemental Information

General Assignments (20%)

Objective: To reinforce theoretical concepts and develop analytical skills through weekly assignments including case studies, research discussions, problem sets, and design tasks directly related to module content.

Labs/Hands-On Activities (25%)

Objective: To provide students with practical understanding of data Science for AI through guided implementation exercises, coding workshops, and interactive demonstrations using industry-standard tools and frameworks.

Midterm Project (20%)

Objective: Individually or in groups, students will complete a project that integrates and applies concepts from the first half of the course. The project requires proficiency in designing and implementing a targeted solution to a specific data science problem or use case relevant to AI.

Capstone Project (25%)

Objective: To provide students with comprehensive hands-on experience in developing a practical, real-world application. Students will integrate multiple NLP techniques covered throughout the course to create an end-to-end solution, complete with detailed documentation and a final presentation.

Course Portfolio (10%)

Objective: Individual dynamic document continuously updated and maintained on GitHub, serving as a comprehensive record of learning progress, code repositories, reflections, and achievements throughout the course. This portfolio demonstrates professional development and technical growth**.**

NAAIC

Course Outline

Week starting on / Module	Module Topic
Week 1	Module 01: The Evolution of Data Science Across AI Paradigms
Week 2	Module 02: AI Data Taxonomy, Sources & Infrastructure
Week 3	Module 03: Data Preprocessing for Machine Learning
Week 4	Module 04: Data Preprocessing for Deep Learning
Week 5	Module 05: Data Preprocessing for Generative AI
Week 6	Module 06: Feature Engineering in ML
Week 7	Module 07: Word Embeddings
Week 8	Module 08: Evaluating and Optimizing ML Models
Week 9	Module 09: Advanced Neural Network Data Handling
Week 10	Module 10: Enterprise Data Pipelines for AI
Week 11	Module 11: Cloud Platforms for AI Implementation
Week 12	Module 12: Data Science for Agentic AI
Week 13	Module 13: Ethical and Responsible AI Data Practices
Week 14	Module 14: Ethical and Responsible AI Data Practices
Week 15	Module 15: Course Wrap up and Review
Week 16	Module 16: Final Exam / Project Due

Supplemental Books (Optional)

Anand, G., & Sharma, R. (2024). *Data science fundamentals & practical approaches*. [Publisher not specified].
Recommended for Modules 1, 3, 6, and 8 for foundational data science concepts and feature engineering.

Ferle, M. (2025). *Snowflake Data Engineer*. Manning Publications.
Recommended for Modules 10 and 11 for cloud-based data pipelines and Snowflake-specific implementations in enterprise AI systems.

Ganesan, K. (2025). *AI and data science for business leaders: A practical guide to implementation*. [Publisher not specified].
Recommended for Modules 10, 11, and 14 for enterprise AI and ethical considerations.

Pierson, L. (2025). *The data science handbook: A practical guide to data analysis and machine learning*. John Wiley & Sons.
Recommended for Modules 3, 4, 6, and 8 for preprocessing and machine learning techniques.

Trummer, I. (2025). *Data analysis with LLMs: Text, tables, images, and sound*. Manning Publications.
Recommended for Modules 4, 5, 7, 12, and 13, with mini-projects for labs (e.g., synthetic data generation) and ethical guidelines for Module 14.

Supplemental Papers (Optional)

Chen, L., & Patel, S. (2024). *Synthetic data generation for robotics using generative adversarial networks*. *IEEE Transactions on Robotics, 40*(3), 1234–1248. https://doi.org/10.1109/TRO.2024.1234567.
Recommended for Module 13 for synthetic data generation techniques in robotics, supporting lab activities with GANs.

Gupta, A., & Nguyen, T. (2025). *Ethical considerations in large language model-driven data science workflows*. *Journal of Artificial Intelligence Ethics, 5*(1), 45–62. https://doi.org/10.1007/s12345-025-67890-1.
Recommended for Module 14 for ethical frameworks in LLM-based data science, enhancing discussions on bias and privacy.

Rodriguez, M., & Kim, J. (2024). *Optimizing scalable data pipelines for deep learning in enterprise AI systems*. *Proceedings of the 2024 International Conference on Machine Learning, 1*, 567–580. https://arxiv.org/abs/2401.12345.
Recommended for Modules 9 and 10 for scalable deep learning pipelines, supporting enterprise AI implementations.

Instructors may consider adopting these books and papers to complement curated resources, aligning with course objectives and teaching style.