

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

AISE 3010 – Data Engineering and Machine Learning

Course Outline Winter 2025

COURSE DESCRIPTION: The aim of this introductory course is to provide a solid background in the fundamentals of data engineering and machine learning. The course covers 1) Introduction to data engineering methods include processing streaming data, data pipelines with CUDA C, big data capability on google cloud platform (GCP), etc. 2) involves recent state-of-the-art deep networks and its learning algorithms for applications related to pattern recognition, feature extraction, image processing, and data argumentation.

ACADEMIC CALENDAR:

https://www.westerncalendar.uwo.ca/Courses.cfm?CourseAcadCalendarID=MAIN_030385_1&SelectCalendar=Live&ArchiveID=

The course covers: 1) Introduction to data pipelines, distributed data management, and streamline data processing; 2) Data manipulation and data structure for big data; and 3) Design and implementation of an engineering group project illustrating the machine learning and data engineering concepts being taught.

PRE OR COREQUISITES: Undergraduate courses in (or SE 2205A/B if taken prior to 2024-25), AISE 2251A/B (or the former SE 2251A/B), AISE 3309A/B (or SE 3309A/B if taken prior to 2024-25), Data Science 3000A/B.

CEAB ACADEMIC UNITS: Engineering Science 100%

CONTACT HOURS:

Timetable information is available at <https://draftmyschedule.uwo.ca/>.

Lectures occur weekly. Laboratory sessions occur weekly.

LECTURE:	3 hours weekly
----------	----------------

LAB:	2 hours weekly
-------------	----------------

RECOMMENDED/REQUIRED TEXT:

1. Adi Wijaya, Data Engineering with Google Cloud Platform, Packt Publishing, 2022. (eBook, \$9.99 at <https://www.packtpub.com/en-us/product/data-engineering-with-google-cloud-platform-9781800561328>)
2. I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016. (freebook at <https://www.deeplearningbook.org>)

RECOMMENDED/ REQUIRED SOFTWARE:

1. Google Cloud Platform
2. Google Colab
3. Pytorch Tutorial: <https://pytorch.org/tutorials/>

RECOMMENDED RESOURCES/REFERENCES:

1. CUDA by example: An introduction to general-purpose GPU programming, Wesley, 2008. (\$62.49 at [Amazon](#))

GENERAL LEARNING OBJECTIVES (CEAB GRADUATE ATTRIBUTES)

Knowledge Base	D	Engineering Tools	D	Impact on Society	
Problem Analysis	D	Individual & Teamwork	D	Ethics and Equity	D
Investigation		Communication		Economics and Project Mgmt.	
Design		Professionalism		Life-Long Learning	

Notation: x represents the content level code as defined by the CEAB. blank = not applicable; I = introduced (introductory); D = developed (intermediate) and A = applied (advanced).

Rating: I – The instructor will introduce the topic at the level required. It is not necessary for the student to have seen the material before. D – There may be a reminder or review, but the student is expected to have seen and been tested on the material before taking the course. A – It is expected that the student can apply the knowledge without prompting (e. g. no review).

COURSE MATERIALS: Weekly content and guides for the laboratories will be available on the course OWL site. The material for this course will be taught in both lectures and labs; therefore, it is imperative that you attend each lecture and lab.

UNITS: SI

COURSE TOPICS AND SPECIFIC LEARNING OUTCOMES:

The following table summarizes the course learning outcomes along with CEAB GAls where the GAls in bold indicate ones to be measured and reported annually.

Course Topics and Specific Learning Outcomes	CEAB Graduate Attributes Indicators
1. Shadow network and its machine learning algorithms At the end of this section, students will be able to: a. understand fundamental concepts of neural network b. describe in detail the specific shadow networks taught in class. c. understand back propagation method. d. understand regularization theory. e. write a program to implement algorithms taught in class for pattern recognition problems.	ET2 KB1
2. Data engineering with GCP At the end of this section, students will be able to: a. understand fundamentals of data engineering. b. understand big data capabilities on GCP. c. describe how to develop a data warehouse with GCP tools. d. understand how to leverage pre-built GCP AI models. e. understand how to process streaming data with dataflow. f. write a program to build machine learning solutions on GCP.	PA2 KB3 ET2
3. Data engineering with basic CUDA At the end of this section, students will be able to: a. understand fundamentals of CUDA architecture in GPU environment for data engineering b. understand CUDA memory management.	PA2 KB3 ET2 ITW2

c. describe how to conduct kernel execution and thread programming to build machine learning solutions for parallel computing. d. understand how to use CUDA streams. e. write a low-level machine learning program using the CUDA NVTX toolkit to achieve acceleration in GPU environment.	
4. Training a convolutional neural network with GCP At the end of this section, students will be able to: a. identify broad categories of convolutional neural network for variety tasks. a. understand concepts of transfer learning, and training from scratch. b. understand how to apply them to process big datasets with GCP. c. write a program using GCP in autoML to train an AI model. d. understand its limitation related to awareness of the principles of equity.	EE4 ET2 KB3 PA2 ITW2

EVALUATION:

Name	% Worth	CEAB GAs ASSESSED
Labs (Total = 4)	20%	ET2, KB1, PA2, ITW2
Final Project (Total = 1)	15%	ITW2, ET2, PA2, KB3
Mid-Term Examination	20%	KB1, KB3, PA2, EE4, ET2
Final Examination	45%	KB1, KB3, PA2, EE4, ET2

Marks will be assigned on the basis of method of analysis and presentation, correctness of solution, clarity and neatness.

COURSE POLICIES:

All work submitted must be of professional quality in the requested format. Material that is handed in dirty, illegible, disorganized, or in an unapproved format will be returned to the student for resubmission and the late submission penalty will take effect. An additional penalty of 10% may be deducted for poor grammar, incoherence, or lack of flow in the written reports.

LABORATORIES: There are 4 laboratory assignments. Some assignments will be individual while some will be assigned in group (up to two). One report can be submitted by each group. Most of the assignments will involve programming in Python or MATLAB. All assignments will be distributed via OWL. All assignments are expected to be submitted via OWL by 11:55 pm on the due date. Each assignment is worth 5% of your overall grade.

FINAL PROJECT: There is one final project which will combine two to three assignments to generate a project to deepen the understanding of our course knowledge. The project could be completed in group or individual. Several options will be provided. Student could pick one of them in order to consider the actual computation resource allocation.

MITERM TEST: There will be one midterm exam that you are keeping up with the material being taught. They will take place during the normal lecture hour.

FINAL EXAMINATION: The final exam will take place during the regular examination period. The final exam will be three hours long, closed book. Only simple, nonprogrammable calculators are allowed.

LATE SUBMISSION POLICY:

Please note that the assignment submission deadline includes flexibility in the form of a 48-hour submission window (grace period). As a result, the instructor reserves the right to deny any requests for academic consideration for assignments submitted after this grace period.

If students submit their assignments beyond the 48-hour grace period, a penalty of 10% per day will be applied for late submissions, up to a maximum of 3 days. After three days, late submissions will no longer be accepted.

ATTENDANCE: Any student who, in the opinion of the instructor, is absent too frequently from class, laboratory periods will be reported to the Dean (after due warning has been given). On the recommendation of the department, and with the permission of the Dean, the student will be debarred from taking the regular final examination in the course.