

Course Introduction

MMA 865 – Big Data Analytics



Smith
SCHOOL OF BUSINESS

Queen's
University

- MMA 865 has two core components:
 - Natural Language Processing
 - Big Data Engineering
- There are two team assignments (30% and 20%) and one individual assignment (30%). *You MUST achieve a passing grade on individual component to pass the course.*
- 20% marks are for Class Participation (more on this later).
- The entire course is designed in Python exclusively. You must complete your assignments in Python.

1. Natural Language Processing

Week 1	Week 2	Week 3	Week 4
<ul style="list-style-type: none">- Course Introduction- What is NLP- Information Extraction- Document Classification- Sentiment Analysis- Machine Translation- Chatbots	<ul style="list-style-type: none">- Text preprocessing- Text vectorization- Topic Modeling	<ul style="list-style-type: none">- Generative AI- Transfer Learning- Large Language Models	<ul style="list-style-type: none">- Deep dive LLM- Fine-tuning- RAG- Use-cases

2 . Big Data Engineering

Week 5	Week 6
<ul style="list-style-type: none">- Overview of Big Data- Hadoop- Spark- Cloud	<ul style="list-style-type: none">- MLOps- ML system design- APIs- Docker and Kubernetes- Data Stores

Assessments

Item	Value	Due*
Class Participation	20%	On-going
Individual Assignment	30%	January 29, 2025
Team Assignment	30%	February 19, 2025
Team Project	20%	February 26, 2025

** By 11:59pm Eastern Time.*

Team Assignment – The Turing Tussle

Your Task

Each team is required to select one of the given topics (below) for the debate. The window for submitting your topic request opens on Day 1 of the class. Submission of your topic and preferred side (*for* or *against*) must be made through email. Ensure you keep at least two topics in mind in case your first choice does not have a competing team.

Each team will then nominate up to three members to represent their team. You *do not* need to submit their names in advance. The format on the day of debate will be as follows:

Activity	Time
<i>For</i> team opening statement	5 m
<i>Against</i> team opening statement	5 m
Interlude	5 m
<i>For</i> rebuttal	5 m
<i>Against</i> rebuttal	5 m
<i>For</i> closing statement	2 m
<i>Against</i> closing statement	2 m
Judge comments	3 m

1. The Impact of Generative AI on Creative Industries

For: Generative AI enhances creativity and innovation in the creative industries.

Against: Generative AI poses a threat to jobs and the authenticity of creative works.

2. The Implications of Artificial General Intelligence (AGI)

For: AGI has the potential to solve major global challenges and benefit humanity.

Against: AGI poses significant risks and could potentially harm humanity.

3. The Use of AI in Surveillance and Privacy

For: AI in surveillance enhances security and public safety.

Against: AI in surveillance infringes on privacy and civil liberties.

4. Should Canada Invest in Building Largescale AI Datacenters for Generative AI?

For: Investing in AI datacenters will boost Canada's technology sector and economy.

Against: The financial costs and other economic factors outweigh the potential benefits.

5. Artificial General Intelligence (AGI) Will Be Achieved by 2027

For: Rapid advancements in AI research and technology will lead to AGI by 2027.

Against: The complexity and other considerations make achieving AGI by 2027 unlikely.

Your Task

For this assignment, you will source the dataset from [House of Commons transcripts database](#). These transcripts provide a rich source of textual data, capturing the full range of parliamentary discussions, debates, and statements.

Your task is to identify a compelling use-case from this dataset, develop a quick proof of concept (POC) to demonstrate the feasibility of your idea, and document your findings and process in a report.

This exercise mirrors real-world scenarios where data scientists and AI practitioners must often navigate ambiguous situations, generate hypotheses, and validate them through data exploration and modeling.

Examples of Use-Cases:

1. **Text Analytics:** Analyze the frequency and context of specific terms or phrases to understand the prominence of certain issues or topics within parliamentary discussions.
2. **Sentiment Analysis:** Assess the sentiment expressed in the debates and speeches, identifying whether discussions on specific topics are generally positive, negative, or neutral.
3. **Generative AI Summarization:** Use generative AI models to create concise summaries of long parliamentary sessions, making it easier to digest and understand the key points discussed.
4. **RAG-Based Chatbots:** Develop a chatbot that leverages retrieval-augmented generation (RAG) to provide accurate and contextually relevant answers to questions about parliamentary proceedings.

HANSARD

JOURNALS

COMMITTEE

BOIE

Detail

List


Search Options: Save Clear all

Results: 1 - 15 of 113320

Hansard - 335

2024-06-19

Show both languages



Lib. (QC)


Hon. Greg Fergus (Hull—Aylmer)

2024-06-19 14:01 [p.25223]

It being Wednesday, we will now have the singing of the national anthem led by the hon. member for Miramichi—Grand Lake.

[Members sang the national anthem]

fX🎵📄📎★



Lib. (NS)

Mike Kelloway (Cape Breton—Canso)

2024-06-19 14:01 [p.25223]

Mr. Speaker, it is a privilege to rise in the House today to recognize a constituent of mine who has received the Prime Minister's Award for Teaching Excellence. Margaret Burke began her career as an early childhood educator in 1978 and has been caring for little ones in her community ever since.

Margaret trail-blazed to create accessible and accommodating learning spaces for children of all abilities. This was particularly impactful during the early years of her work, as during this period, many complex learning needs went

...More

fX🎵📄📎★

PARLIAMENT AND SESSION

Date Range

All Debates

44th Parliament

🕒

1st 2021.11.22 -

43rd Parliament

2nd 2020.09.23 - 2021.08.15

1st 2019.12.05 - 2020.08.18

42nd Parliament

1st 2015.12.03 - 2019.09.11

More...

ORDER OF BUSINESS

Individual Assignment

PART 2

- Identify one use-case of Artificial Intelligence at your work. It can be a conventional NLP use-case like sentiment analysis, document classification, information extraction, etc. or it could be use-case that leverage Generative AI and Large Language Models (LLM) like ChatGPT, Bard, Midjourney etc.
- The identified use-case **must** use text data or any other unstructured data like images, audio, video, etc.
- Supervised or unsupervised learning use-cases on structured data (numeric and categorical features only) are **NOT** acceptable for this assignment.
- The use-case may involve ad hoc analysis or a proposed AI functionality for your company's product.

Instructions

- Write a document in 1000 words or less. **You must clearly organize your document in following Sections:**

- **Section 1:**
 - Define the problem statement.
 - What will happen if you solve this problem and what will happen if you maintain status quo.
- **Section 2:**
 - Explain how you are proposing to solve this problem with AI techniques.
 - Go deep into the solution workflow without using too much technical jargon.
- **Section 3:**
 - Quantify the impact of the solution.
 - Document your assumptions for quantification.
- **Section 4:**
 - Create a high-level plan to implement this use-case.
 - Give details about how you will collect the data, what kind of data is required, what kind of preprocessing you think is required.
 - What type of modeling is required.
 - How you will measure and select the model.

Assessment Summary

- Class Participation 20% **(On-going)**
- Individual Assignment 30% **January 29, 2025**
- Team Assignment 30% **February 19, 2025**
- Team Project 20% **February 26, 2025**

Questions on Deliverables?

- **Transformers for Natural Language Processing:** Build, train, and fine-tune deep neural network architectures for NLP with Python, PyTorch, TensorFlow, BERT, and GPT-3, **2nd Edition**
- **Learning Spark: Lightning-Fast Data Analytics, 2nd Edition**
- **Modern Generative AI with ChatGPT and OpenAI Models 1st Edition** - Leverage the capabilities of OpenAI's LLM for productivity and innovation with GPT3 and GPT4

Office Hours

- **TA (Raghav Gupta):**
 - Online every **Thursday 5:00-6:00 pm ET**
 - Zoom link for office hours on calendar
 - Email: Raghav.Gupta@queensu.ca
- Meeting with professor can be done by appointment only. Send email for appointment moez.ali@queensu.ca

By end of this course

1. **Understand Natural Language Processing (NLP):** Gain foundational knowledge of NLP and its key use cases, such as information extraction, document classification, sentiment analysis, and topic modeling.
2. **Implement Classical NLP Use Cases:** Build end-to-end NLP pipelines and implement classical NLP use-cases in Python.
3. **Explore Generative AI and Modern NLP:** Learn about Generative AI, Large Language Models (LLMs), Transformer architectures, and their applications in NLP.
4. **Machine Learning Operations (MLOps):** Understand the principles of MLOps, including API design, Docker containerization, ML system monitoring, Kubernetes and more.
5. **Big Data Concepts:** Dive into Hadoop, Spark architectures, and the differences between SQL and NoSQL data storage solutions.
6. **Develop Hands-On Applications:** Build a full-stack project combining Python back-end development, external APIs, data processing pipelines, NoSQL database, and minimalistic front-end design with HTML (with help of AI).

SmithWriter Blog Collection

Best Christmas Gifts in 2024

The Ultimate Guide to Finding the Perfect Presents This Holiday Season

[Gift Guides](#)

Christmas 2024 gift ideas holiday shopping

Winter Fashion in Canada: Embrace the Cold with Style

Your ultimate guide to staying warm and fashionable during Canada's winter months

[Fashion](#)

winter fashion Canada cold weather style winter outfits

Top Travel Destinations in 2025

Explore the Must-Visit Places for Your Future Travels

[Travel](#)

travel 2025 holiday destinations travel trends

Changing Baby Diapers: A Comprehensive Guide for New Parents

Everything You Need to Know to Make Diaper Changes Easy and Stress-Free

[Parenting](#)

baby care diapers parenting tips newborn

Dealing with Stupid Managers: A Guide for Employees

Strategies to Navigate the Challenges of Ineffective Leadership

[Career Development](#)

management challenges employee support career advice office dynamics

Learning Chess at 50: A Guide to Mastering the Game Later in Life

Unleashing Your Inner Strategist No Matter Your Age

[Chess, Hobbies, Self-Improvement](#)

chess learning age 50 strategy mental health

Moving from Canada to the USA for a Better Life: Your Comprehensive Guide

Everything You Need to Know Before Making the Big Move

[Relocation](#)

migration expat advice life in America

The Future of Artificial Intelligence in Healthcare

Exploring Innovations and Transformations in Patient Care

[Health Technology](#)

Artificial Intelligence Healthcare Innovations Future of Medicine Patient Care

Top 10 Python Libraries for Data Science

Explore the Essential Tools for Data Analysis and Machine Learning

[Data Science](#)

Python Data Science Machine Learning Data Analysis Data Visualization

Understanding Blockchain Technology for Beginners

A Simple Guide to Grasp the Concepts and Applications of Blockchain

How to Start a Successful Freelance Career

A Comprehensive Guide to Launching Your Freelance Business

The Importance of Mental Health in the Workplace

Why Prioritizing Mental Wellness is Essential for Productivity and Well-being

Best Christmas Gifts in 2024

The Ultimate Guide to Finding the Perfect Presents This Holiday Season

Christmas 2024 gift ideas holiday shopping

Introduction

The holiday season is upon us, and with it comes the age-old challenge of finding the perfect gifts for our loved ones. As we approach Christmas 2024, it's crucial to stay ahead of the trends and discover meaningful presents that will bring joy to friends and family. This guide is designed to help you navigate the myriad of options available, ensuring you find gifts that cater to unique interests and lifestyles. From the latest tech gadgets to heartfelt personalized items, we will cover diverse categories to make your holiday shopping experience enjoyable and successful.

Trending Gift Categories for 2024

This year, certain categories are rising in popularity, making them the go-to choices for thoughtful gift-giving. Understanding these trends can help guide your selections and make your gifts stand out.

Tech Gadgets

In 2024, tech gadgets continue to dominate the gift landscape. From smart home devices to innovative wearables, there's something for every tech enthusiast. Consider the Apple AirPods Pro for superior sound experience, or smart assistants like the Amazon Echo Show for a blend of utility and fun. These gadgets not only enhance daily activities but also appeal to the growing desire for integrated technology in our lives.

Eco-Friendly Products

There has been a noticeable shift towards sustainability, with eco-friendly products gaining traction as perfect gifts for the conscious consumer. Products made from recycled materials, such as bamboo kitchenware or biodegradable phone cases, make thoughtful presents that align with environmental values. Brands like Klean Kanteen offer stylish reusable drinkware that reduces plastic waste while being practical.

Top Picks for Each Recipient

Finding the right gift often depends on understanding the recipient's personality and preferences. Here's a curated list of top picks across various categories to inspire your selection.

Gifts for Him

For the men in your life, consider gifts that blend practicality with luxury. Tech gadgets such as the latest

Using GPT-4o-mini can produced 1700 blog / 24 hour for less than \$5/day cost

Example 2

Liberal Party Platform Explorer

Analysis

The Liberal party's position on taxes indicates they are planning to increase certain taxes, specifically targeting the wealthiest individuals and the largest corporations that have been highly profitable. They plan to raise corporate taxes on banks and insurance companies earning over \$1 billion annually, and to introduce a temporary Canada Recovery Dividend from these firms as they have recovered quickly from the impacts of COVID-19. Additionally, they aim to ensure that high earners pay at least a minimum tax of 15% to eliminate loopholes that allow them to avoid paying taxes through deductions and credits. There will also be new taxes on luxury items like high-end vehicles and boats. The overall goal is to make the tax system fairer by asking those who can afford it to contribute a bit more, which they believe will help in rebuilding the economy and supporting public services for all Canadians.

Related Documents

Similarity: 79.6%

Page 192

Forward. For Everyone. 187 Modernized the tax system to recognize the growing importance of digital tech giants and made

Similarity: 78.4%

Page 40

Forward. For Everyone. 35 • Require landlords to disclose, on their tax filing, the rent they receive pre- and post-renovation

Similarity: 77.3%

Page 41

Forward. For Everyone. 36 Stop Excessive Profits in the Financialization of Housing Large corporate owners of residential

It can help 18 million Canadians next year to decide who should they vote for

Class Participation

Base	Additions	Reductions
Each student starts with a base score of 14 points out of 20. This is calculated as 2 points per class session for a total of 7 sessions.	Answering questions asked by the Professor: +2 point for each correct or thoughtful answer provided during class.	Joining late: First time: -1 point Second time: -2 points Third time and beyond: -5 points per occurrence
	Participation in class activities: +2 point for actively engaging in any class activities or group work.	Leave early to catch train -5 for first instance. More than 1 time = 0 on class participation

MMA865 - Big Data Analytics

Cloning the Repository and Installing Dependencies

1. Install Anaconda:

- Windows:

- a. Download the Anaconda installer from the [official Anaconda website](#).
- b. Run the installer and follow the instructions.

- macOS:

- a. Download the Anaconda installer for macOS from the [official Anaconda website](#).
- b. Open the downloaded .pkg file and follow the installation instructions.

- Linux:

- a. Download the Anaconda installer for Linux from the [official Anaconda website](#).
- b. Open Terminal and navigate to the directory where the installer was downloaded.
- c. Run the installer using the command:

```
bash Anaconda3-<version>-Linux-x86_64.sh
```



Let's get started
