



**Lingnan University
Hong Kong**

**Detecting and Visualizing Echo Chambers in Reddit Vaccine Myths
Discussions**

CLE9035 Final Project Proposal

Course Name: Information, Misinformation, and the Media

Professor: Dr. Jen Foster

Student Name: BEGIMBAEVA Alima Nur

SID: 4072257

PART A – OVERVIEW (20 points)

[A1] The nature of the project

The proposed project for the CLE9035 class targets to find and analyze existing echo chambers' presence in Reddit discussion based on vaccine myths. The project will use an available dataset from Kaggle that includes posts and comments from the r/VaccineMyths subreddit. Subsequently, it will detect and illustrate the communities that omit specific information and discredit other sources. By applying existing data analysis methods (like sentiment and topic analysis, network-based community detection) on the posts and comments, it will get a basic understanding of misinformation surrounding vaccination in social networks.

[A2] The methodology

The project will use the available dataset on Kaggle and combine it with programming language python using tools (like pandas, NetworkX) to construct networks and connecting users through their posts and comments. Using some algorithms (like Louvain and Girvan-Newman) will aim to demonstrate densely located users that inevitably form echo chambers. Contents of the text of the posts and comments will go through some processes (like topic modeling through genism LDA and sentiments analysis with VADER/TextBlob) to show the characterization of the groups. The findings will subsequently be illustrated through graphs. It is expected to take around a month.

[A3] The goals

The goals of the given project are to show the utilization of data science tools to perform tasks and connect it to philosophical ideologies such as echo chambers. It will aim to examine the roles of groups or communities of an agent in shaping their beliefs surrounding pandemics, vaccination and medicine. It also seeks to judge and shed light on the effects of echo chambers on the public's understanding and beliefs of vaccines to share strategies on combatting the epidemic of online misinformation.

PART B – TARGET CONCEPTS (10 points)

- Echo Chambers
- Misinformation
- Sentiment Analysis

PART C – EXPLANATION (20 points)

• [C1.1] Echo Chambers Justification

I believe that the formation and presence of echo chambers is extremely dangerous and not widespread. Echo chambers lead even agents in good faith to form not epistemically good or correct beliefs, which is of great danger for our society. Which is why this project chooses to focus on the given topic.

- **[C1.2] Echo Chambers Plan**

The project is designed to demonstrate the presence of echo chambers in the daily social lives of people using reddit communities, to demonstrate how widespread the phenomenon is and its dangers.

- **[C2.1] Misinformation Justification**

Misinformation, especially medical misinformation, is a critical topic that literally costs people their lives. The spread of misinformation, especially online in the last few decades has been reaching its boiling point, making it an essential topic to cover in this study connecting epistemology and data science.

- **[C2.2] Misinformation Plan**

By examining discourse in the r/VaccineMyths subreddit, this project aims to raise awareness about the misinformation spread in online communities, and how simple it is to fall for these traps.

- **[C3.1] Sentiment Analysis Justification**

Sentiment analysis is an essential analytical method on quantifying the emotions and feeling of texts, especially in cases such as comments and posts, which are found everyone in social networks. The method of quantifying something so unquantifiable is very intriguing and will be able to demonstrate the data more clearly to the public.

- **[C3.2] Sentiment Analysis Plan**

The project will use the sentiment analysis to show how positive or negative comments from the subreddit form will be within the echo chamber and how it actually influences agent beliefs and interactions. It is a critical tool to probe online public opinion and the relationship of misinformation.