



**UTM**  
UNIVERSITI TEKNOLOGI MALAYSIA

**SCHOOL OF COMPUTING**  
Faculty of Engineering

Project Proposal Form MCSD 6215  
Sem:...1..... Session:.....

## SECTION A: Project Information.

---

Program Name: **Masters of Science (Data Science)**

Subject Name: **Project 1 (MCSD 6215)**

Student Name: Alexander Tan Ka Jin

Metric Number: MCST1

Student Email & Phone: tanka@graduate.utm.my 016-7778810

Project Title: Simulation and Analysis of how TikTok users break free from information cocoons

Supervisor 1: \_\_\_\_\_

Supervisor 2 / Industry  
Advisor(if any): \_\_\_\_\_

## SECTION B: Project Proposal

---

### Introduction:

Misinformation is a prevalent effect on social media sites. Often, scientific misinformation and political misinformation is spread through various channels and sites such as X (formerly Twitter), Facebook, Youtube and Tiktok. Engagement with such videos and content from prominent superspreaders of misinformation may lead to recommender algorithms suggesting more content from the same or similar superspreaders a phenomenon known as information cocoons. This proposal focuses on TikTok, a platform that allows users to quickly engage in short, minute long videos and how engagement in misinformative videos leads to recommender algorithms sculpting a personalized environment that reinforces more radical forms of misinformation.

### Problem Background:

Recommender Algorithms are algorithms that chooses information and media that suits the user's preferences. Most popular social media sites like Reddit, Youtube and Instagram utilize these algorithms to allow users to see content that they may prefer with the benefit of increasing user engagement with the platform by filtering out media that users may not prefer. Recommender algorithms are divided into two types, content-based filtering and collaborative filtering. Content-based filtering identifies the media that a user has previously engaged with, and find media that is like those media. Collaborative filtering identifies what a user engages with and suggest content based on other people that share similar interests. These algorithms have long incorporated machine learning in their identification of user patterns. Both algorithms eventually form what are known as information cocoons which are personalized digital environments that

appeal only to user interests and limit the diversity of information and alternative views on a given subject that the user experiences. In 2020, the Covid-19 pandemic reached a critical mass and started spreading internationally. In medical institutions, medical doctors scrambled to contain the virus and shared information on potential cures and vaccines for the virus. In social media however, users wrestled and debated over the factuality of the information from news sources of various backgrounds such as the rumor that hydroxychloroquine could cure the virus. Many of these misinformation come from dedicated sources spreading conspiracy theories like InfoWars (whose ownership is soon to be transferred to satirical news site The Onion) and often lead to other sources that spread similar “alternative facts” about the world. Combined with skepticism towards mainstream media outlets makes conspiracy communities susceptible to these cocoons.

**Problem Statement:**

User engagement with these pieces of misinformation like the hydroxychloroquine myth may result in recommender algorithms picking up these patterns and suggesting media that discusses similar views of the fake treatment. This will result in the formation of an information cocoon that entraps users into engaging with even more media that affirms existing misinformed viewpoints. Studies on this issue analyze how diverse the content is provided by these recommender algorithms through random-walks and analysis of user behaviors collected over the years like the study by Lei Hou et al. in 2023 that analyzed recommender algorithms through random walk or an analysis of user behavior by Nian Li et al. in 2022 on the Chinese social media platform KuaiShou, a short form video platform that is similar to TikTok. Both studies analyzed how cocoons are formed from the beginning of a user’s interaction with the social media platform but the data of how a user breaks free from the cocoon is still understudied. Given that there are some diversity maintained in digital recommendations after being sculpted by recommender algorithms, there exists a that heavily personalized digital environments may still allow people to reach media that provide more accurate information than misinformation superspreaders.

**Aim of the Project:**

The project will provide an analysis to what extent can groups of users break out of information cocoons via the use of simulated random walks starting from a digital environment that is heavily personalized.

**Objectives of the Project:**

1. Review existing literature on information cocoons, information spread on social media and the use of recommender systems in social media
2. To create several simulated random walks starting from an already personalized environment and collect data regarding the short-form videos they engaged with.
3. Visualize the timeline in the diversity of videos and the degree in which these random walks have successfully filtered out the content that was present in the start of the walk.
4. Draw conclusions based on findings and suggest research overcoming the project’s limitations.

**Scopes of the Project:**

The research looks at how recommender algorithms are affected by established use within short-form video content platforms like TikTok. Its focus on the conspiracy community within these platforms is due to skepticism of mainstream media within conspiracy communities causes a more insular community prone to such information cocoons. The research's use of TikTok is due to it being a relatively new social media platform and its short-video content allowing more finer data points on how recommender algorithms evolve. The simulated random walks do not fully simulate psychological motivations for a user's choice in content but are analyses of the extent recommender algorithms promote diverse content that challenges a user's interest.

**Expected Contribution of the Project:**

The research would contribute to the study of how information is spread online and the study of information cocoons. It would hopefully motivate social media platforms to change recommender systems such that it allows users to discover content that diversifies their opinions and interests.

**Project Requirements:**

Software:	Python, TikTok Web API
Hardware:	Basic Computing Hardware
Technology/Technique/ Methodology/Algorithm:	Random walk

**Type of Project (Focusing on Data Science):**

- ☐ Data Preparation and Modeling
- ☒ Data Analysis and Visualization
- ☐ Business Intelligence and Analytics
- ☒ Machine Learning and Prediction
- ☐ Data Science Application in Business Domain

**Status of Project:**


- ☒ New
- ☐ Continued

If continued, what is the previous title?

I declare that this project is proposed by:

[ ☒ ] Myself

[ ☐ ] Supervisor/Industry Advisor ( )

  
Signature

Date .....

The Supervisor(s) shall complete this section.

I/We agree to become the supervisor(s) for this student under aforesaid proposed title.

Signature \_\_\_\_\_ Date \_\_\_\_\_

Signature \_\_\_\_\_ Date \_\_\_\_\_

---

The Evaluator(s) shall complete this section.

**Result:**

<input type="checkbox"/> FULL APPROVAL	<input type="checkbox"/> CONDITIONAL APPROVAL (Major)*
<input type="checkbox"/> CONDITIONAL APPROVAL (Minor)	<input type="checkbox"/> FAIL*

\* Student has to submit new proposal form considering the evaluators' comments.

Name of Evaluator 1:

Signature

.....  
Date

Name of Evaluator 2:

Signature

\*\*\*\*\*  
Date