CAPSTONE PROJECT
IDEA GENERATION

OPTION #1 – CORRELATING CRYPTOCURRENCY PRICE WITH TWITTER SENTIMENT ANALYSIS

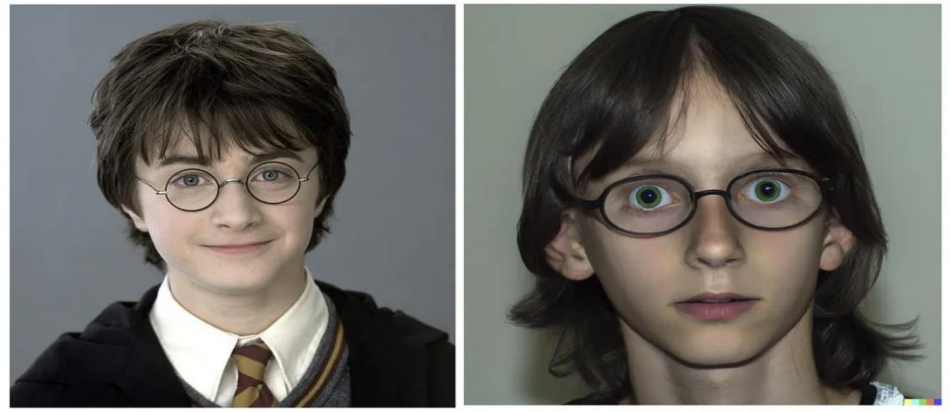
- Background
  - Research has shown that real time tweets has become a popular way to gain public sentiment and based on this assumption you can predict the direction of the prices of different cryptocurrencies.
- Objective
  - Compare across different classification models and their effectiveness in predicting:
    - price deflation (0)
    - price stagnation (1)
    - price hikes (2)
- Data to be scraped from Twitter
  - Text from tweets and top level comments scraped from twitter and use python libraries to obtain a "polarity" score.
  - "retweet" and "like" numerical indicators are combined with polarity score to represent sentiment
  - "price" will be represented by the closing prices of the different cryptocurrencies from yahoo finance.

OPTION #2- THE OPPOSITE OF DALLE

- Background
  - OpenAI's Dall-E2: generates images from textual descriptions
  - Generate textual descriptions from images
    - Instagram caption generator
    - Meme caption generator
- Objective
  - Ride on the burgeoning global Instagram influencer market
  - "Laughter is the best medicine." - meme caption generator
- Data
  - Scraping Instagram using a profile crawler powered by Selenium
  - Stock Market memes from Twitter – select a specific theme to narrow the scope

Harry Potter

Book description: A boy with thin face, black hair and bright-green eyes. He wore round glasses. very thin scar on forehead



Movie **Dall-E2**

OPTION #3- DIABETES STAGE PREDICTIVE CLASSIFICATION MODEL

- Background
  - Diabetes is one of leading death causes in Singapore.
  - Leveraging on AI to predict patients' diabetes based on medical measurements and patients' personal hobbies could help efficiently screen and filter for patients with diabetes and predict diabetes stage in large scale to enhance productivity.
- Objective
  - Develop a machine learning/deep learning classification model to predict the diabetes stage of a patient (*Diabetes_012*)
  - Design the experiment for training/validation/testing and evaluate the model performance comprehensively.
  - Give reasonable interpretations on the classification result, both locally (e.g., For a particular patient, why the prediction result is 0/1/2) and globally (e.g., Which features are most influential to the classification results).
- Data
  - A *diabetes_health_indicators.csv* file with 253680 patients' information.
  - The target variable Diabetes_012 has 3 classes. 0 is for no diabetes or only during pregnancy, 1 is for prediabetes, and 2 is for diabetes. This dataset has 21 feature variables.
  - Need to seek clearance to use it for personal project – dataset provided by IHIS.

METHODOLOGY