# CS765 ASSIGNMENT - 3

**Abhishek Dixit (22m0761)**
**Rishabh Singh (22m0762)**
**Osim Abes (22m0825)**

---

# Introduction

This section involves the basic description of the features of our DApp, which are necessary to explain the handling of the mentioned issues in Part A of the assignment.

1. Attributes of Voter
    a. Unique user ID
    b. A dictionary of topics mapped to their rating in the form of **Elo rating** (https://en.wikipedia.org/wiki/Elo_rating_system )

2. Attributes of News Item
    a. Unique ID
    b. Title - The heading for the specific news item
    c. Content
    d. Author
    e. Topic - The domain of the news item, i.e., politics, sports, health, etc.
    f. isRated - Has the news item been rated yet or not?

# Part A

## (1) Sybil attack:

We assign weights to votes based on the **Elo rating** of the voter; this gives more influence to voters with higher credibility. This means that contributions from trusted and experienced voters carry more weight in determining the truthfulness of a news article.
We also introduce a committee comprising the top 1% of the total voters, selected from the top 10% of the highest-rated voters. This adds an extra layer of security and reliability to the fact-checking process.

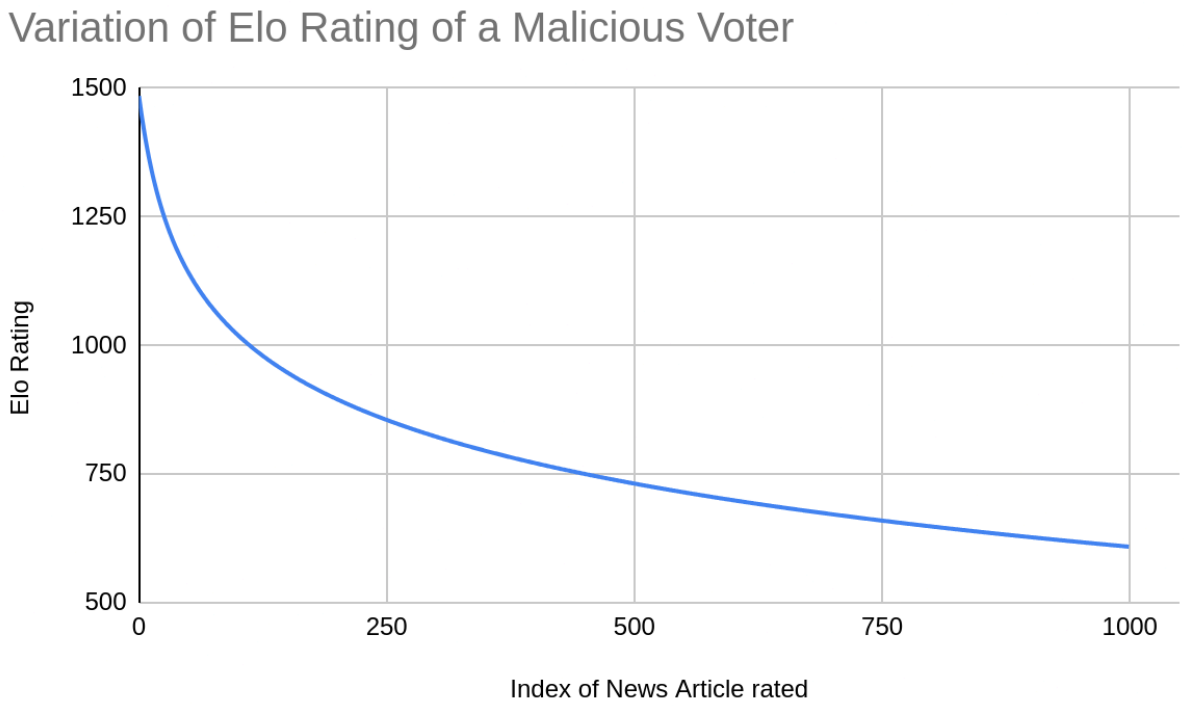
Even if a malicious voter manages to create multiple fake accounts to influence the voting process, their impact is limited because the committee determines the change in the Elo rating of all voters; hence, a voter cannot affect his own Elo rating maliciously, and his rating falls very quickly otherwise.
Also, by nominating the top 10% of the highest-rated voters and selecting committee members from volunteers within this group, the system ensures that only individuals genuinely committed to the integrity of the fact-checking process participate in the committee.

$$E_B = \frac{1}{1 + 10^{(R_A - R_B)/400}}$$

Where $E_B$ is the expected score of the voter and $R_A$ is a baseline value which we assume to be 1500

## (2) Method to evaluate or re-evaluate the trustworthiness of voters:

We evaluate the consistency of a voter's ratings with the weighted average rating determined by the committee for each news article. A voter's trustworthiness can be assessed based on how closely their ratings align with the consensus opinion of the committee.
Voters whose ratings consistently align closely with the committee's consensus should have their trustworthiness rating increased, while those whose ratings deviate significantly should have their rating decreased. Elo Rating calculation works such that higher-rated people suffer more penalties for making mistakes, and lower-rated voters have a chance to climb up the ladder quickly.
To prevent gaming of the system, we've implemented a safeguard where a voter's rating dynamically adjusts based on their voting behavior relative to the committee's average rating. If a voter consistently gives accurate assessments and earns a position in the committee but later begins to act maliciously by deviating significantly from the committee's consensus, their rating will decrease accordingly. This ensures that the system continually assesses and adjusts the trustworthiness of voters and the committee, making it difficult for individuals to manipulate the system for their gain.



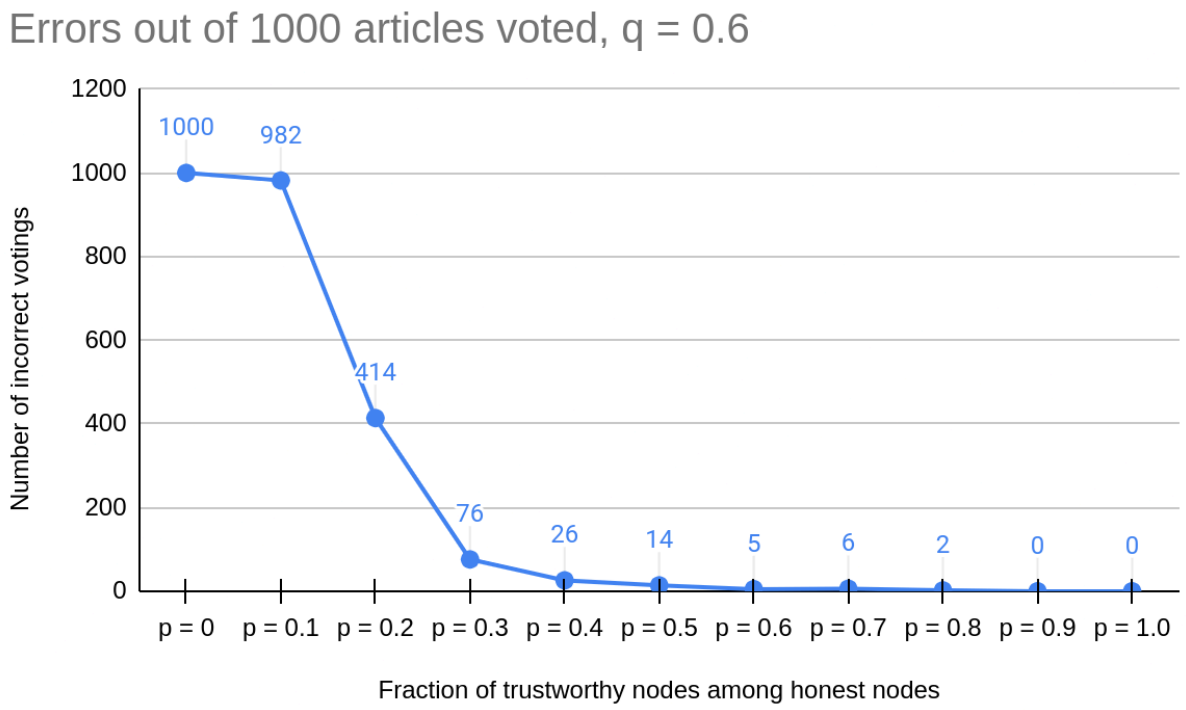

## (3) The opinions of more trustworthy voters should be given more weight

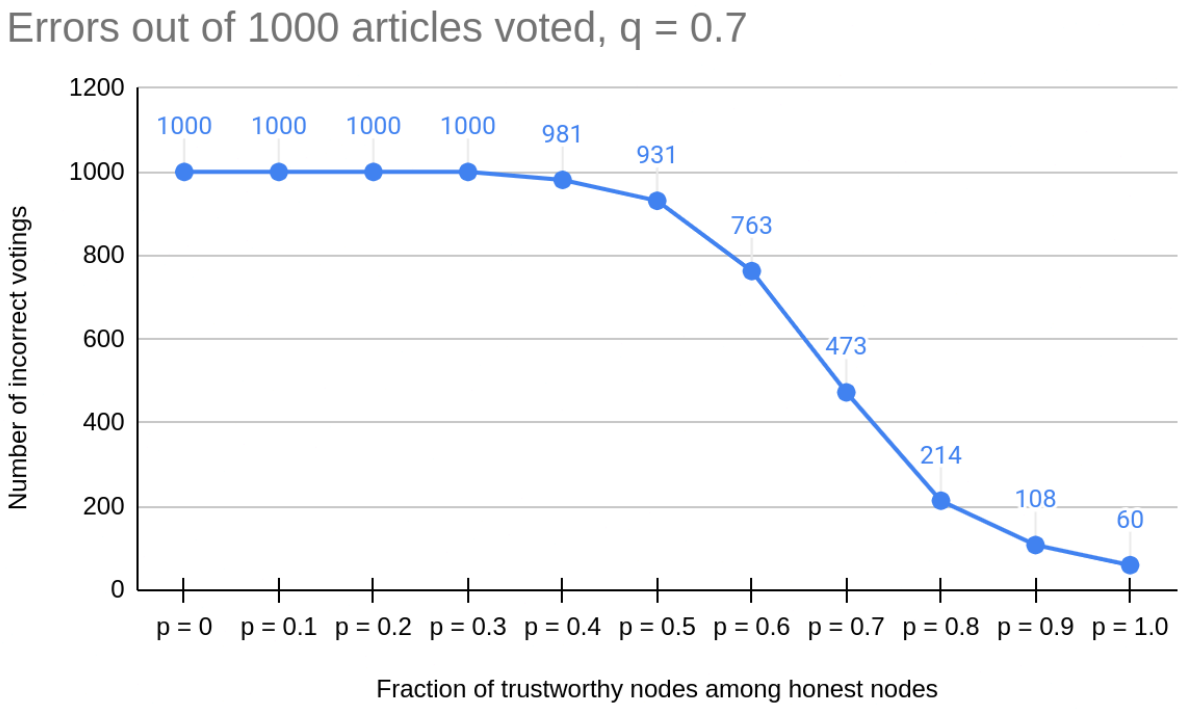
To address the issue of varying trustworthiness across different topics, we utilize a topic attribute for each news item and maintain Elo ratings for each voter topic-wise. This enables us to form committees of voters who excel in specific topics. Implementing topic-wise ratings for voters adds a layer of sophistication to the fact-checking process, enhancing the accuracy and reliability of assessments across diverse subject matters. This approach acknowledges that trustworthiness may vary depending on the subject matter, allowing for a more nuanced evaluation process. Whether it's physics, politics, economics, or any other topic, the system dynamically forms committees comprising the most highly rated voters in that particular area, ensuring that evaluations are conducted by those with the deepest understanding and proficiency. This focused evaluation not only mitigates the risk of misinformation but also fosters a community of specialized experts committed to combating fake news across a wide range of domains.

The formula used for the final public rating of a news article:

R = (Sum(Elo_rating[voter]*news_rating_given)/Sum(Elo_rating[voter]))

It can be clearly observed that a person having a higher rating has a higher weightage in determining the final rating.






## (4) Rational voters are to be incentivized

To encourage rational and truthful participation from voters, a system of incentives is implemented where voters are rewarded for accurately reflecting their knowledge and judgment in their votes. By aligning their votes with what they believe to be true, voters increase the likelihood of their assessments closely matching the consensus reached by the committee. As a result, their Elo rating, indicative of their trustworthiness, gradually rises. This increase in rating signifies their commitment to honest participation. It enhances their chances of being selected to join the committee, where they can contribute to shaping the collective decision-making process. In this way, the system incentivizes voters to engage in the fact-checking process with integrity, fostering a culture of truthful participation and contributing to the overall reliability of the DApp.

We didn't include any monetary incentives, despite it seeming to be an obvious inclusion because monetary incentives can cause voters to give ratings in a way that can get them maximum reward instead of voting honestly to the truest of their ability and the purpose of the app of being a fact checker quickly changes into that of being a betting hub. Also, monetary benefits can make people try and form malicious groups to skew results in their favor to earn more money quickly.

## (5) Uploading a news item

To identify a news item in the Dapp, we do content hashing. Content hashing provides an efficient method for identifying news items within the DApp. By subjecting the textual content of each news article to a hashing algorithm, a unique hash value is generated, serving as a distinct identifier for the article. This hash allows for quick and accurate retrieval of the news item from the DApp's database, regardless of size or complexity. Utilizing content hashing ensures data integrity and security, as even minor alterations to the article's content would result in a completely different hash value. Thus, content hashing offers a reliable and efficient means of identifying news items within the DApp, facilitating seamless access and evaluation for fact-checking purposes.

We didn’t use the title or author exclusively for identifying as they can be similar for multiple different articles.

If Identifying the problem statement refers to the identification of the topic of the article, we can apply Machine Learning algorithms involving NLP, which is beyond the scope of this assignment.

## (6) Bootstrapping:

A systematic approach is required to bootstrap the fact-checking process, otherwise the DApp would lack trustworthy ratings for initial voters and might evolve unsuitably from there on. One effective method is to initiate an examination of users' trust values for each topic by requesting them to rate a set number of articles per topic (3 here). These articles are randomly selected from a continuously updated database initially populated with admin-rated articles and subsequently enriched with user-rated articles. By evaluating users' ratings of these articles, an initial trust rating can be assigned based on their performance in the examination. This process allows for the gradual accumulation of trustworthy ratings, enabling the DApp to establish a reliable foundation for fact-checking news items. Over time, as more users participate and

rate articles, the accuracy and depth of the trust ratings will improve, further enhancing the effectiveness of the fact-checking mechanism within the DApp.

---