Problem:

A news value maximiser: Politically and commercially affiliated media companies are tasked with maximizing the views for certain articles more than others. Build a system that maximises the views for these “aligned” articles.

The problem can be modelled as a K-arm Bandit problem, where each arm corresponds to an article. The aim is to maximize the views of "aligned" articles and ensure that other articles are also being considered to gather sufficient data.

Arms: Each article is an arm in the K-arm Bandit.

Reward: The number of views an article receives after it is shown to users.

Aim: Maximize the views for aligned articles, but also explore other articles to avoid bias and gather data.

Initialization:

Define a set of articles ‘A’ and identify the subset of aligned articles.

Exploration:

Greedy: With probability ϵ, explore a random article from the entire set A. Otherwise, exploit by selecting the article with the highest expected reward (views). For aligned articles, assign higher initial expectations or give them priority in exploitation.

Reward Collection:

Track the number of views each article gets after being selected.

For aligned articles, we can introduce a slight boost to bias the system subtly towards these articles.

Update Strategy:

Update the estimated rewards for each article based on the number of views it receives.

Balance Exploration and Exploitation:

Ensure that the system occasionally explores non-aligned articles to gather data and avoid complete bias as it might lead to suboptimal outcomes.

Evaluation:

Monitor the total number of views and the proportion of views on aligned articles. Adjust the exploration strategy parameters (like ϵ) based on the results.