# ASSIGNMENT

#### SHARAT KN

In cancer treatment clinical trials, researchers often face uncertainty regarding the effectiveness of different treatment options for various types and stages of cancer. Similarly, the multi-armed bandit problem deals with uncertainty about the reward distributions associated with different actions (arms). The relationship between cancer treatment clinical trials and the multi-armed bandit problem lies in their decision-making processes under uncertainty and exploration-exploitation trade-offs. In cancer treatment clinical trials, researchers often face the challenge of efficiently allocating resources (such as budget and time) to different treatment options while considering various factors like treatment success rates, patient characteristics, and resource constraints.

## Code Walkthrough:

##### Environment Setup:

* + We have defined a Open AI gym environment representing the cancer treatment decision-making process, encapsulating the state space, action space, and reward mechanism. The environment allows the agent to select treatment options and provide feedback in the form of rewards.
  + When an action is selected, we filter all the rows in the dataset containing that particular treatment type. Then we select a random row from that filtered dataset

##### Algorithm Implementation:

* + We have implemented a Multi-Armed Bandit algorithm, epsilon-greedy, to learn the optimal treatment selection strategy. The algorithm balances exploration (trying new treatments) and exploitation (using known effective treatments) based on a specified exploration parameter (epsilon).
  + We set epsilon to 0.1
  + We assign Rewards based on the success rate of the treatment. We also give weightage to budget, and time considerations. We also take into account if the tumour was malignant or benign, Malignant tumour holding higher reward.
  + Return or Cumulative Reward - combines the rewards using a weighted sum, with weights chosen based on their relative importance.

##### Training Loop:

* + We then conduct training episodes where the MAB agent interacts with the environment, selecting actions, observing rewards, and updating its policy based on feedback. During each episode, the agent collects information about treatment effectiveness and refines its decision-making strategy.

##### Evaluation:

* + After training, we evaluate the learned policy by examining the Q-values associated with each treatment option.
  + We then assess the convergence of Q-values over training episodes and analyze the balance between exploration and exploitation achieved by the algorithm.

##### Analysis and Interpretation:

* + The learned policy is then analyzed to identify the most effective treatment options based on the highest Q-values.

##### Conclusion and Recommendations:

* + We finally conclude the study by summarizing the findings and recommending the most promising treatment options for further evaluation in clinical trials. In our case, we have found Hormone Therapy to be the most effective treatment.