# AI INNOVATION 3.0

24-Hour-Hackathon

27 APRIL 2024 – 28 APRIL 2024

## PROBLEM STATEMENTS

YAMAHA MOTOR SOLUTIONS INDIA

# Problem Statement 1

## *DocBot (Chatbot based document assistance system)*

The DocBot challenge aims to develop an intelligent chatbot-based assistance system capable of efficiently processing documents containing both text and images. Participants are required to create a user-friendly interface allowing users to input documents and interact with the chatbot to seek information related to the document content. Upon receiving queries, the chatbot should retrieve or generate meaningful textual answers along with the most relevant image(s) from the document, specifying the corresponding page number(s) where they are located.

**Prerequisites:**

1. Deep Learning Skills
2. Chatbot Development
3. Document Processing
4. Web Development
5. Familiarity with Information Retrieval Techniques: Understanding information retrieval methods such as keyword-based searching, document indexing, and relevance ranking will facilitate efficient retrieval of relevant content from the input documents.

**Important Information:**

1. Participants are permitted to utilize any open services or libraries available for the task, including pre-trained models or APIs for NLP, computer vision, and chatbot development.

2. Emphasis should be placed on designing an intuitive and user-friendly interface to enhance the user experience while ensuring the accuracy and efficiency of the DocBot in retrieving relevant information from documents.

3. Additional Marks: Participants can earn bonus marks if the DocBot is capable of assisting users in the Japanese language as well.

**Dataset:**

The dataset containing sample documents (pdfs) will be provided to participants.

**Evaluation:**

This problem won't contain any specific evaluation metric; instead, submissions will be assessed based on various criteria:

1. User-Friendly Interface: The ease of use and intuitiveness of the interface will be evaluated to ensure a seamless user experience.
2. Efficiency of Chatbot: The efficiency and effectiveness of the chatbot in retrieving accurate and relevant answers from the input documents will be assessed.

3. Accuracy of Images Retrieval: The accuracy of retrieved images along with their corresponding page numbers will be scrutinized to ensure precise information retrieval.
4. Multilingualism: Bonus marks will be awarded for incorporating multi-lingual capabilities into the DocBot, particularly if it can assist users in Japanese.

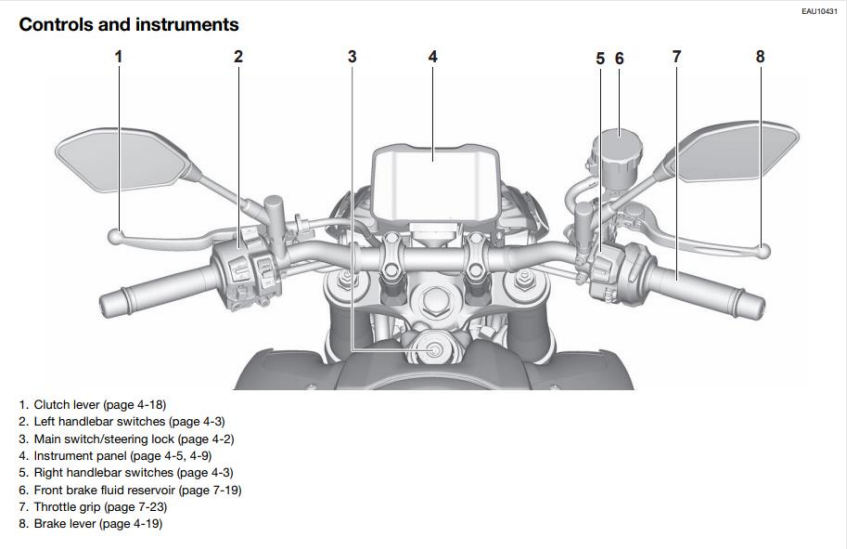
Participants are encouraged to focus on these aspects while developing their solutions to ensure a well-rounded and effective DocBot system.

**Sample question prompt and expected answer from DocBot:**

Prompt: Where is the clutch Lever of bike?

Reply:

The clutch lever is located at the left-hand side of the steering. Refer to page 13.



# Problem Statement 2

## *HistologyNet: Self-Supervised Segmentation Solver (S4)*

**HistologyNet** aims to revolutionize biomedical image segmentation by harnessing the power of self-supervised learning. Traditional annotation methods for histology images are time-consuming and costly. In this challenge, participants will develop a self-supervised learning model for efficient binary segmentation of histology images, without the need for extensive manual annotation.

**Task:**

1. Develop a self-supervised learning model capable of segmenting histology images into meaningful regions without extensive manual annotation.
2. Choose and implement an appropriate pre-text task to facilitate self-supervised learning.
3. In addition to developing the segmentation model, participants must create a user-friendly webpage. This webpage will allow users to input histology images and obtain their segmented masks, demonstrating the practical utility of the developed models.

**Dataset:**

Participants will be provided with a training dataset of histology images, comprising both labelled and unlabelled images.

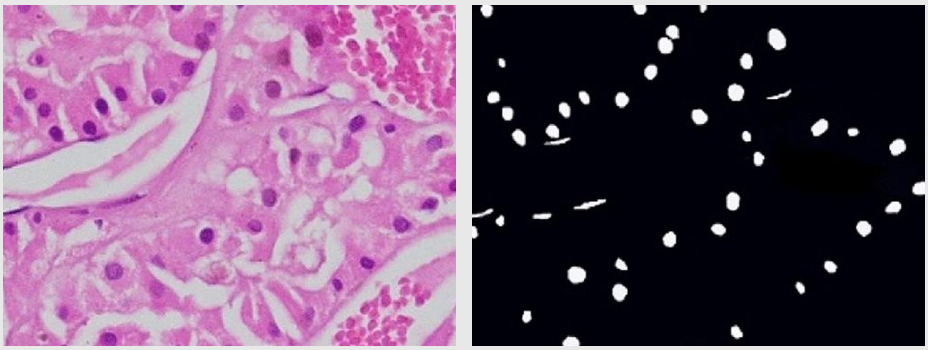
**Evaluation:**

The performance of participants' models will be evaluated based on two key metrics:

- Jaccard Index
- Dice Coefficient

For webpage, the ease of use, intuitiveness and efficiency of the interface will be evaluated to ensure a seamless user experience.

**Sample Histology Image-Ground Truth Pair:**



# Problem Statement 3

## *VineNet: Grasp of the Grapes*

Viticulture is the science and art of grapes cultivation. In the bustling world of viticulture, precision is paramount. **VineNet: Grasp of the Grapes** is an exciting hackathon challenge tailored to revolutionize vineyard management through cutting-edge deep learning techniques. The task at hand? Crafting a lightweight fast deep learning model to precisely segment grape bunches in high-resolution vineyard images.

**Task:**

1. Participants are tasked with building a deep learning model capable of accurately segmenting grape bunches (semantic segmentation) in high-resolution images of vineyards.

2. Emphasize minimizing inference time to enable real-time applications in vineyard management.

3. As an integral component, participants must develop a user-friendly webpage showcasing the practical application of their segmentation models.

**Dataset:**

A meticulously curated training dataset containing labelled high-resolution images of grape vineyards will be provided to participants.

**Evaluation Metrics:**

Models will be evaluated based on the following performance metrics:

1. Mean Average Precision (mAP)

2. Precision

3. Recall

4. Inference Time

**Sample grape vineyard image along with ground truth overlaid on the image:**

# Problem Statement 4

## *ClassyPics - Class-Conditional Image Generation*

Image generation is a cutting-edge application of deep learning, especially beneficial when training data is limited. In this hackathon challenge, participants will tackle the task of generating images of specific objects using conditional stable diffusion models.

**Task:**

Participants are tasked with designing and implementing a conditional stable diffusion model capable of generating synthetic data samples corresponding to specific class labels. The main objectives include:

1. Training the model effectively.
2. Developing a user interface that accepts class labels and produces corresponding images.

**Prerequisites:**

1. Solid understanding of Deep Learning Fundamentals.
2. Proficiency in Python and Deep Learning Frameworks.
3. Knowledge of Probability and Statistics.
4. Experience with Image Processing and Computer Vision.

**Considerations:**

Participants must implement mechanisms for class-conditional generation, enabling the model to produce synthetic data samples tailored to specific class labels.

**Datasets:**

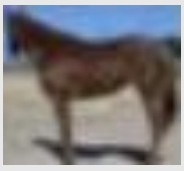
Two different class-labelled datasets will be provided to the participants.

**Evaluation:**

1. Quality of Generated Images: Evaluation will involve a subjective assessment of the quality of the generated images, considering factors like clarity, realism, and relevance to the specified class label.
2. Additional Evaluation Metrics: Detailed evaluation metrics will be shared soon to provide participants with clear guidelines for assessing the performance of their models.
3. Efficiency of User Interface: The efficiency and usability of the user interface developed for generating images will be evaluated, focusing on factors such as responsiveness, ease of use, and overall user experience.

**Sample**

Prompt: *Horse* → Output:



# Problem Statement 5

## *ImageSpeak: Empowering Visual Understanding*

ImageSpeak is a hackathon challenge aimed at enhancing accessibility for individuals who are blind or visually impaired. The challenge focuses on developing a deep learning model capable of generating descriptive captions for images, enabling users to gain meaningful insights from visual content.

**Tasks:**

1. Develop a deep learning model capable of generating descriptive captions for images taken by individuals who are blind or visually impaired. The model should accurately describe the content of the images in a concise and informative manner.

2. Integrate the developed model into a mobile application, allowing users to capture images and receive descriptive captions in real-time.

**Dataset:**

The dataset comprises images, each paired with five descriptive captions. Participants will utilize this dataset to train and evaluate their deep learning models for image captioning.

**Additional Information:**

Participants will earn additional marks if they incorporate speech output functionality into the application. Utilizing text-to-speech APIs or libraries is allowed for this task.

**Evaluation:**

1. Quality of Captions: Evaluation will involve a subjective assessment of the quality of the captions, considering factors such as clarity, coherence, and relevance.
2. Evaluation Metrics: Detailed evaluation metrics will be provided to participants, offering clear guidelines for assessing the performance of their models.
3. Efficiency of User Interface: The efficiency and usability of the mobile application will be evaluated, focusing on factors such as responsiveness, intuitiveness, and overall user experience.

**Sample Images with their captions:**

A computer screen with a Windows message about Microsoft license terms.

A can of green beans is sitting on a counter in a kitchen.

A photo taken from a residential street in front of some homes with a stormy sky above.

A blue sky with fluffy clouds, taken from a car while driving on the highway.