# Generative AI - Assignment 3

December 4, 2024

## Question 1

Pick up a diffusion model from one of the provided ones below:

- stabilityai/stable-diffusion-3.5-medium

- Efficient-Large-Model/Sana_1600M_1024px

- stabilityai/sdxl-turbo

- PixArt-alpha/PixArt-XL-2-1024-MS

- playgroundai/playground-v2.5-1024px-aesthetic

- InstantX/InstantID

- prompthero/openjourney

- ByteDance/Hyper-SD

- dalle-mini/dalle-mini

- stablediffusionapi/anything-v5

- Kandinsky-2

Let me know which model you are using for record keeping purposes.
   Use it to

- Generate an image that contains 10 hands. Yes, in a single image.

- Generate an image where two people are shaking hands.

- Generate an image where 5 people are performing a handshake simultaneously.

Try to get to the desired solution as close as possible, and document your findings. Submit your code as well.

## Question 2

Fine-tune the model that you have just picked in the previous question on a dataset that specializes in hands. The code is already prepared over here and a sample dataset is also provided for your convenience.
   Report your results and document the findings. See how it improved, and find reasons as to why it is behaving in this manner.

# Question 3

This is Neil Harbisson.



This man is considered by The World™as the first ever **cyborg**. I wish I was joking.

An antenna is embedded within his brain. What that antenna does, apart from a lot of things, is to let him hear colours and see sounds.

Yes, the notion itself is absurd, but there is a phenomena for it - **Synesthesia**.

Your job is to construct a workflow where you simulate the idea of Synesthesia.

- You will take audio as input. This audio can be anything.

- You will pass it through a workflow that you will design.

- At the end of the workflow, your diffusion model from Question 1 lies. It will use whatever has been processed by your workflow so far to construct a `perception`, an idea of what the audio "looks" like.

TL;DR - you will input audio and you will get a visual depiction of what the audio "looks" like.

Here are some guidelines for this question:

- You can use, apart from your previous diffusion model, ANY OTHER MODEL OR MODELS that may be useful for your workflow.

- That means you can pick any audio-based or language-based model that may assist you in the problem.

- **If any two groups use the same diffusion model, there will be consequences for both groups.** Please make sure you are using a model NO ONE ELSE IS. This applies for all questions.

Here's what you have to submit:

- Questions 1, 2, and 3 will be executed on the Kaggle Platform. We cannot provide you any access to servers with GPUs in CureMD due to security reasons. Don't worry, the **Nvidia Tesla P100** graphics card should be more than enough for what you're trying to do.

- You will submit code and documentation for questions 1 and 2.

- For question 3, you will submit your entire workflow/architecture and a short guide on how it works and what it does.

- Submit a LaTeXdocument for all questions. **I will not entertain excuses for not being able to write a LaTeXdocument this time**. You will have to use (and is preferred for you to use) TeXStudio for making LaTeXdocuments.

Keep in mind that I will conduct your viva for this assignment.
Some general information noted here:

- You can only work in teams of **two** people. No more, no less.

- The submission for all of this stuff is before **13th December 2024, 6:59 PM** at **abdullah.faiz@curemd.com**. Make sure to add Aun Abdullah and Naeem Hassan in CC. While I am lenient about most things, **I will not tolerate submissions after the stipulated deadline.** It's better you submit something than nothing at all.

- *Try not to plagiarise your work.* This is meant for your learning, and there wil be no point in learning if you can't tell me the reasons for why you've done something in the assignment.

Lecture slides are shared with you all. In case of any queries, you should approach me even when I'm busy.