**Multi-model AI with Transformer: A Tutorial**

**4th Nepal AI School, Kathmandu**

24th May 2023

In this tutorial, we will play with a pretrained transformer as a hands-on introduction to the multi-modal AI using transformers. This tutorial is designed to help you to set up and run a multi-modal transformer, and explore the given model’s strengths and weaknesses. In the later lectures, you will explore the principles behind designing and implementing such models.

**Chapter I: Experimental setup**

1. Download the source code and model to your colab from:

<https://drive.google.com/drive/folders/1wtVhvwjT-lQRikTHo3Gb8bYb2u24VyVi?usp=share_link>

1. Run the provide source code for different tasks provided to you

**Chapter II: Exploration**

1. Run the provide model on your own image using the following syntax

your\_img = load\_image\_from\_url('link\_to\_your\_image')

1. Execute and list out different task from the following paper:

# *Unified-IO: A Unified Model for Vision, Language, and Multi-Modal Tasks* <https://arxiv.org/pdf/2206.08916>

**Chapter III: Tasks**

1. Make your own image mask and use the following syntax to generate an image.

uio.segmentation\_based\_generation(image\_mask,["object\_class"], num\_decodes=1)

1. Find out which output of the network (if any) could be fed back to the network to recover the original image as closely as possible.

If you have any questions, feel free to write to [danda.paudel@naamii.org.np](mailto:danda.paudel@naamii.org.np).