## Project ID: CS44

**Project Source:** School of Computer Science

**Project Title:** Can transformers solve a preschool image understanding problem?

## Project Description and Scope:

Visual Question Answering (VQA) is an extremely stimulating and challenging research area where Computer Vision (CV) and Natural Language Processing (NLP) have recently converged. In VQA, semantic information in the same media must be compared with the semantics implied by a question expressed in natural language, doubling the artificial intelligence-related effort. This project aims to employ transformers to develop a pictorial question–answer method for addressing questions related to counting ability, color concepts, knowledge of geometric shapes, and the perception of different objects by kids. Further, it is required to explore Language Studies (Native) Games online for kids in Pre-K at tinytap.com. This website has collected numerous pictures, questions, and answers suitable for preschool education. We aim to build a suitable dataset that can be used to evaluate the method.

## Expected outcomes/deliverables:

Method with the source code on GitHub
Research report

## Specific required knowledge, skills, and/or technology:

machine learning and Python

## Fields that this project may involve:

Data Science/Analytics;Artificial Intelligence;NLP;

## Resources provided by the client:

Yes, the client will provide dataset.

Datasets:
https://github.com/skarifahmed/pic2question

https://www.tinytap.com/content/all/ages-4-5/category/104/language-studies-native/

https://github.com/ad-freiburg/large-qa-datasets

Resources:
https://arxiv.org/abs/2103.02937
https://link.springer.com/article/10.1007/s00521-021-06080-w#Sec3