Shruthi.S
311521243049

Final Project

PROJECT TITLE

VISUAL QUESTION & ANSWERING USING TRANSFORMER LIBRARY

AGENDA

1. Introduction
2. Requirements
3. Usage
4. Data Preparation
5. Model Architecture
6. Training
7. Evaluation
8. Deployment
9. Contributing
10. License

PROBLEM STATEMENT

Developing a Visual Question & Answering (VQA) system using the Transformer library to enable computers to understand and answer questions about visual content in images.

PROJECT OVERVIEW

The objective of this project is to leverage the Transformer library, a state-of-the-art NLP toolkit, to build an effective VQA system. By training Transformer-based models on paired images and questions with corresponding answers, the system will learn to understand the content of images and generate accurate responses to questions about them.

WHO ARE THE END USERS?

The end users of the Visual Question & Answering (VQA) system include the general public, educators, medical professionals, customer support personnel, researchers, and content creators seeking assistance or information related to visual content.

YOUR SOLUTION AND ITS VALUE PROPOSITION

The solution is a Visual Question & Answering (VQA) system powered by the Transformer library. It efficiently interprets visual content, providing accurate answers to user questions. Its value proposition lies in enhanced understanding of visual information, improved user engagement, versatility across domains, streamlined workflows, and adoption of innovative technology.

THE WOW IN YOUR SOLUTION

The "wow" factor of this solution lies in its seamless integration of natural language processing and computer vision, enabling computers to understand and respond to questions about visual content with human-like comprehension. This transformative capability, powered by the Transformer library, offers users an engaging and efficient way to interact with visual information across diverse domains.

MODELLING

The solution utilizes Transformer-based architectures, pre-trained models, computer vision techniques, and fine-tuning to build a Visual Question & Answering system. These models understand both textual and visual inputs, fuse them effectively, and generate accurate responses to questions about visual content.

RESULTS

The results show that the Visual Question & Answering (VQA) system achieves high accuracy, demonstrates strong understanding of both textual and visual inputs, exhibits robust performance across diverse content, and contributes to user satisfaction and operational efficiency.

Demo Link