Visual Curiosity Engine: Predicting Human Question-Provoking Regions in Images

***Abstract*—Most vision models focus on saliency prediction, i.e., estimating where humans will look in an image, but curiosity differs from attention. While saliency is primarily driven by contrast, motion, or distinctiveness, curiosity is triggered by ambiguity, novelty, anomalies, or contextual mismatches. This project introduces the new task of *visual curiosity prediction*: identifying regions in an image most likely to provoke human questions. Unlike saliency maps, which capture "where the eye goes," curiosity maps aim to capture "where the mind wonders."**

The ability to predict such curiosity-inducing regions has wide applications, including marketing (highlighting unusual or engaging product features), interactive assistants (flagging intriguing content), and human-AI interfaces (enabling machines to ask clarifying questions about ambiguous elements). The main challenge lies in the subjective nature of curiosity and the absence of a standardized dataset.

To address this, we plan to curate 200–400 images from COCO, Open Images, and VQA datasets, annotated with bounding boxes, curiosity scores, question types (why/what/how/anomaly), and example questions, aggregated into heatmaps and region rankings. Methodologically, we propose a ResNet-50 or Vision Transformer backbone with a U-Net-style decoder for pixel-wise curiosity heatmap prediction, coupled with object-level region ranking using features from pretrained detectors (DETR, Faster R-CNN). Multi-task learning will be employed with focal loss for heatmap prediction, pairwise hinge loss for ranking, and an optional transformer decoder for question generation.

Evaluation will be conducted using Pearson correlation with human heatmaps, NDCG@k for ranking quality, and top-1 overlap accuracy, supplemented by human A/B studies against saliency-based and random baselines.

This work establishes curiosity prediction as a novel computer vision task, contributing the first dataset and baseline models, with potential impact in adaptive education, intuitive human-AI collaboration, and AI systems that better align with human cognition and creativity.