# Eye Tracking - Homework 4
# A Transformer-Based Model for the Prediction of Human Gaze Behavior on Videos

Muhammad Turab

November 9, 2024

## Summary

Eye-tracking technology that uses human gaze to analyze videos is becoming more important. To automate video analysis with eye-tracking data, we need to accurately mimic how people look at videos. However, this is difficult because human gaze patterns are often complex and unclear. This paper introduces a transformer-based approach for the prediction of human gaze behavior in videos. They predict human gaze in third person videos using a reinforcement learning model based on transformer. The model learns to simulate human gaze behavior to predict fixation points. The model uses a pre-trained ResNet for feature extraction and the Decision Transformer for sequential modeling to improve gaze prediction accuracy compared to other methods. To validate the accuracy of this method a gaze-based action recognition dataset was used from VirtualHome where this model outperforms existing models in terms of both precision and angular error.

The key findings of this work show that the proposed model can effectively mimic human gaze behavior with a significant improvement in downstream tasks such as action recognition and anticipation. The model achieves promising results in activity recognition by using the predicted gaze instead of true gaze which shows the potential application of this model where actual gaze data is limited. The study concludes that the integration of reinforcement learning with a transformer offers a robust solution for automated gaze prediction in video analysis and understanding [ÖRA+24].

## References

[ÖRA+24] Süleyman Özdel, Yao Rong, Berat Mert Albaba, Yen-Ling Kuo, Xi Wang, and Enkelejda Kasneci. A transformer-based model for the prediction of human gaze behavior on videos. In *Proceedings of the 2024 Symposium on Eye Tracking Research and Applications*, pages 1–6, 2024.