# See Clearly, Act Intelligently: Transformers in Transparent Environments

**Author** Omar Elamin **Supervisor** Jinke He **Responsible Professor** Frans A. Oliehoek

## Background & Motivation

- Transformers have displayed remarkable performance in capturing dependencies in **temporal** data.
- Responsible for the impressive performance of LLMs like **ChatGPT**.
- Reinforcement Learning Environment models learn the **dynamics** of their environment. This is a temporal relationship that transformers have demonstrated notable performance in capturing.
- Previous work has demonstrated improved **sample efficiency** [1] & superior **performance** [2], but very little is known about why and what specifically allows these models to perform so well.
- We therefore control the environment and observe **future prediction** capability to gain more insight.

## What is a Transformer?

- A transformer uses a mechanism called **self-attention** to infer about how various parts of the sequence affect each other [3].

The small brown **fox** | jumped over the lazy dog

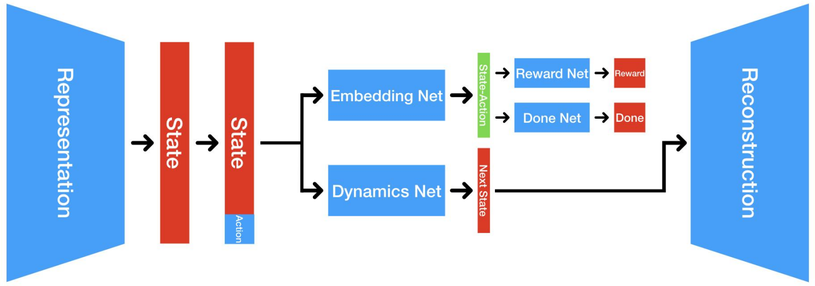
- An **attention head** acting on this sentence might learn to look for adjectives, to gain insight into what the word **fox** means in this contex
- What a head searches for is a learned parameter, and has no drop-off fc long-range dependencies.
- Given **N** attention heads, each one can be run in parallel.
- These two factors are what give transformers an advantage over traditional recurrent methods.

## Research Question

Under what conditions do **transformers** enhance the **planning performance** of model-based reinforcement learning in **fully observable** environments compared to conventional methods?

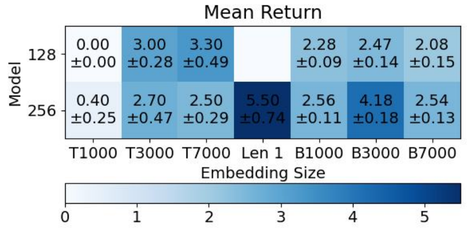
## Methods

- Using the MinAtar breakout environment.
- This is **fully observable** & **Markov**, meaning the past contains no additional information for predicting the next state.
- Evaluation is done using **Monte-Carlo Tree Search**. This involves determining an **expected value** of a state by simulating **multiple steps** into the future using the **world model**.
- This multiple step future prediction is where transformers could excel.
- A deep learning architecture based on DreamerV2 [4] is used as the baseline.
- This architecture involves learning a **latent** representation of the environment, and passing in only the **current state** into a deep neural network, to predict the following state.
- The **transformer** architecture is **identical**, except transformer dynamics network takes the **state-action embedding** (green below) as **input** instead.
- The sequence length parameter for the transformer was fixed to 20.




## Experiments & Results

- To answer the question, we investigated two conditions and their impact on performance: **Model Size** and **Data Availability**.
- We used **128** & **256** embedding dimension to test model size
- The models were also trained on **1000**, **3000**, & **7000** episodes to test data availability. We also ran an experiment with sequence length 1.
- Transformers **failed** to outperform the baseline in nearly every configuration.




- These results present a **limitation in the dataset**. It's inherent structure being episodes of a **DQN** agent learning to play breakout means the earlier episodes mostly consist of **early stage states**.
- This means the models that were given less episodes were also **overfitting** to early game scenarios, which is a possible explanation for the drop in performance for the baseline model when given more data.
- This overfitting is even more **prominent** when the model is **larger**, which could explain the more drastic decrease in performance when there is a larger latent space. Future work should train with **more** data that is more **diverse**.
- The experiment using sequence length 1 performing so well suggests that the **past** information adds **no value** in these **fully observable** environments. It seems to **introduce** noise and additional **error**.

## References

[1] V. Micheli, E. Alonso, and F. Fleuret, "Transformers are sample-efficient world models," in **The Eleventh International Conference on Learning Representations**, 2023.
[2] M. Schwarzer, A. Anand, R. Goel, R. D. Hjelm, A. Courville, and P. Bachman, **Data-efficient reinforcement learning with self-predictive representations**, 2021.
[3] A. Vaswani, N. Shazeer, N. Parmar, et al., **Attention is all you need**, 2017.
[4] D. Hafner, T. Lillicrap, M. Norouzi, and J. Ba, **Mastering atari with discrete world models**, 2022.