Learning How Humans Play Board Games with GPT-4IAR

UNIVERSITY OF HELSINKI
FACULTY OF SCIENCE

Victor Yeom-Song[1], Daisy Lin[3], Ionatan Kuperwajs[3], Heiko Schütt[3], Wei Ji Ma[3], Luigi Acerbi[1,2]

[1]Department of Computer Science, University of Helsinki, Finland [2]Finnish Center for Artificial Intelligence FCAI,

[3]Center for Neural Science and the Department of Psychology, New York University

FCAI Finnish Center for Artificial Intelligence

Main Points

- Board games are a good model to study human planning and decision making
- Learning a good model that predicts gameplay can be a great tool to study the inner mechanisms of planning
- We present a model that shows great promise for this task!

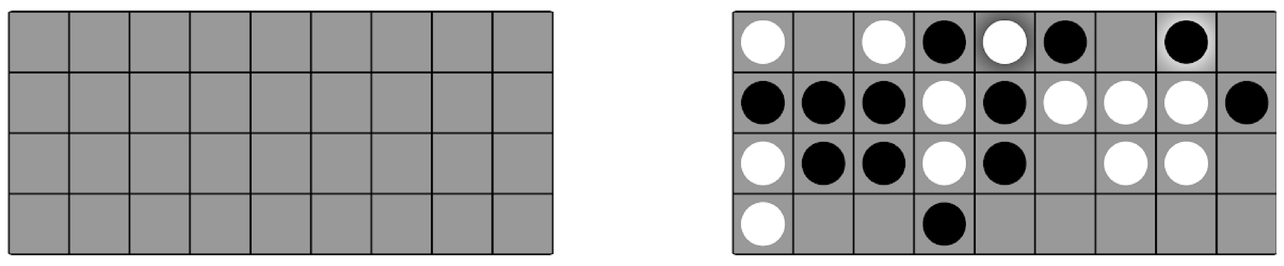
Introduction

Motivation

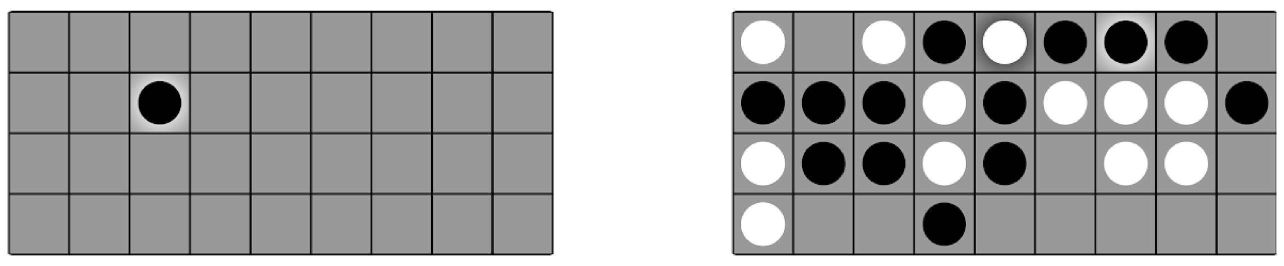
In Cognitive Science, planning and decision making are active areas of research. There is great interest in understanding the inner mechanisms of how expertise affects the way humans plan, and the main focus for this type of study has been placed on games [1].

Game and Task

We study the game known as 4-in-a-row (4IAR), proposed first by van Opheusden et al. [2]. This is a generalization of Tic-Tac-Toe, where players seek to connect 4 pieces of their color in a 4×9 board. Black to move, predict the next action on these boards:



The first board has too much entropy, and on the second the human lapses!



It is *really* hard to learn how a human plays!

This game is simple enough that it is feasible to tractably model (as opposed to e.g. chess), while maintaining enough complexity for it to extract intricacies of how we strategize.

An interpretable model

The first model to study how expertise affects how humans plan was proposed by van Opheusden et al. [3, 2]. This is a hand-crafted cognitive model that seeks to be highly interpretable, which asserts that the planning process follows a decision tree structure whose depth increases with expertise.

Going Further

Deep neural networks are able to capture complex patterns which may have been missed in building the previous model. Thus, a fully connected network was used by Kuperwajs et al. [4] to impose a noise ceiling on the move predictions. This can, in turn, be used to gauge the "goodness" of the model, as well as understand which behaviors the interpretable model struggles to capture, and improve it for the corresponding scenarios. On a dataset with ~10M data points, the network achieves a cross-entropy loss of 1.866 and an accuracy of 41.71 %.

Transformers and Attention

The crux of the Transformer architecture, first proposed by Vaswani et al. [5], is the attention mechanism. As presented in the survey by Lin et al. [6], the architecture has had great success in tasks with sequential data, and recently it has been used to tackle sequential decision making [7, 8, 9]. We propose to use this architecture for 4IAR.

The Data

Currently we use the computational model proposed by van Opheusden et al. [2] to generate synthetic games, and use ~1M data points for training. Inspired by the Trajectory Transformer, presented by Janner et al. [7], we represent a sequence of a board, an action and duration of a move through tokens:

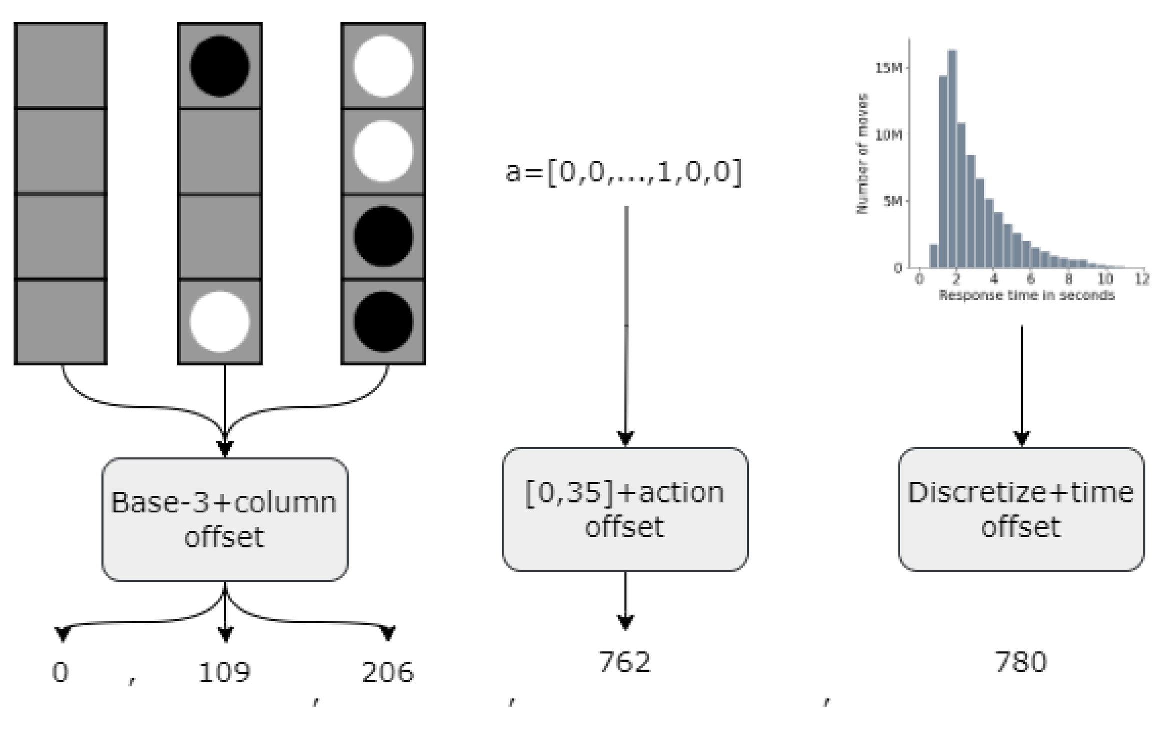


Figure 1. Tokenization scheme of board states, actions and durations.

The Architecture

In a similar fashion as the data, the architecture is influenced by the Trajectory Transformer [7]. A conceptual diagram is shown:

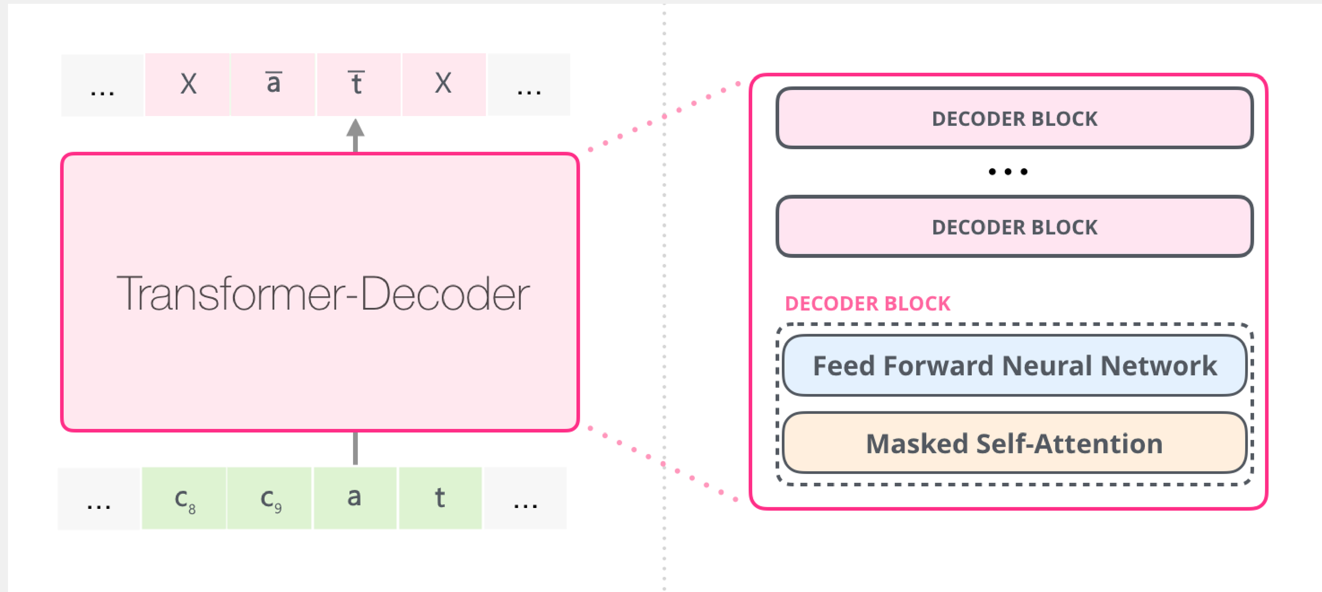


Figure 2. Proposed Neural Network Architecture. Figure adapted from "The Illustrated GPT-2".

We have used a model with approximately 85 million parameters (network weights and biases), which results from using the same hyperparameters as the standard GPT-2 [10] on a smaller vocabulary size. With this setup, we produce the results shown here.

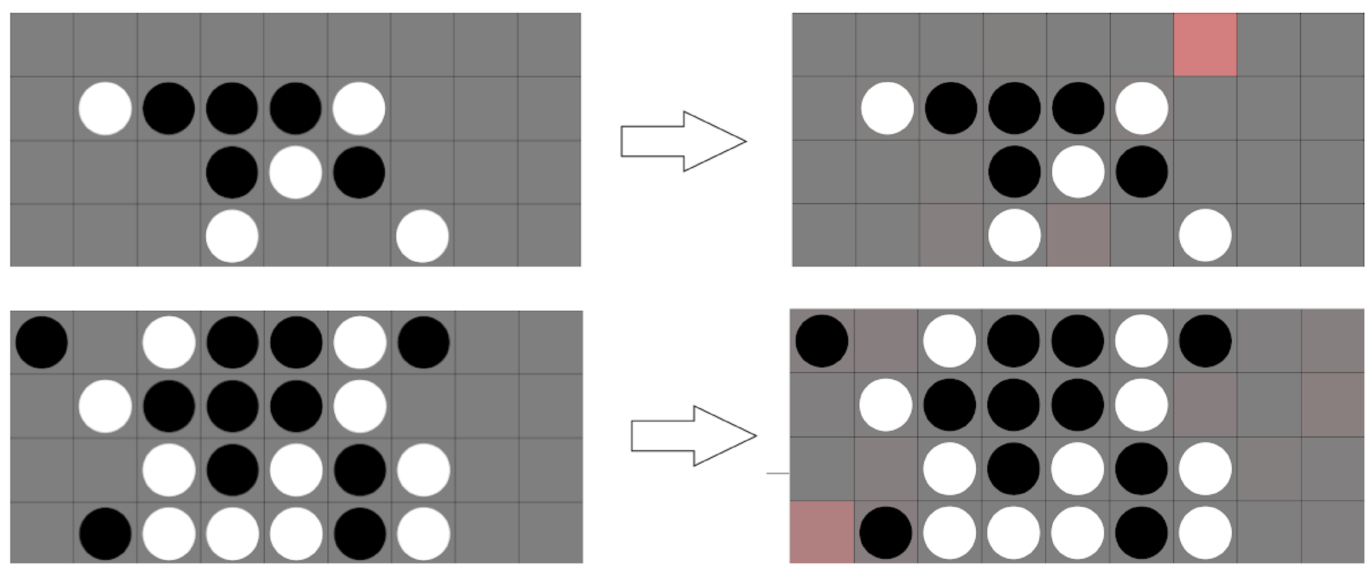
Experiments

Metrics

- Cross-Entropy Loss;
- Prediction accuracy on moves (AccMove) and reaction times (AccRT);
- For reaction times, RMS Error (RMSE).

Some results

With the plan to train this architecture on the ~10M point data set in the future, we present some preliminary results from the network trained on the synthetic set.



Model comparisons

	Cross Entropy Loss	AccMove (%)	AccRT (%)	RMSE (# bins)
Untrained GPT-4IAR	7.19	0.03	0.06	5.66
Trained GPT-4IAR	1.44	45.6	21.3	3.23
Fully Connected	1.866	41.71	—	—

These preliminary results are promising. The network is shown to learn the task well, and shows good signs for being able to predict how long an agent may take to produce a move!

Next Steps

- Train the network on the same data set as Kuperwajs et al. [4].
- Conduct scaling studies, changing the number of parameters used to verify if we are using the optimal hyperparameters.
- Produce comparisons between our architecture and the fully connected network to look for possible temporal dependencies.

References

[1] Kelsey R Allen, Franziska Brändle, Matthew Botvinick, Judith Fan, Samuel J Gershman, alison gopnik, Thomas L Griffiths, Joshua K Hartshorne, Tobias U Hauser, Mark K Ho, and et al. Using games to understand the mind, Feb 2023.

[2] Bas van Opheusden, Ionatan Kuperwajs, Gianni Galbiati, Zahy Bnaya, Yunqi Li, and Wei Ji Ma. Expertise increases planning depth in human gameplay. *Nature*, 618(7967):1000–1005, Jun 2023.

[3] Bas van Opheusden, Gianni Galbiati, Zahy Bnaya, Yunqi Li, and Wei Ji Ma. A computational model for decision tree search. In *CogSci 2017 - Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, CogSci 2017 - Proceedings of the 39th Annual Meeting of the Cognitive Science Society: Computational Foundations of Cognition, pages 1254–1259. The Cognitive Science Society, 2017.

[4] Bas van Opheusden, Gianni Galbiati, Ionatan Kuperwajs, Zahy Bnaya, Yunqi li, and Wei Ji Ma. Revealing the impact of expertise on human planning with a two-player board game, Feb 2021.

[5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[6] Tianyang Lin, Yuxin Wang, Xiangyang Liu, and Xipeng Qiu. A survey of transformers. *AI Open*, 3:111–132, 2022.

[7] Michael Janner, Qiyang Li, and Sergey Levine. Offline reinforcement learning as one big sequence modeling problem. In *Advances in Neural Information Processing Systems*, 2021.

[8] Lili Chen, Kevin Lu, Aravind Rajeswaran, Kimin Lee, Aditya Grover, Misha Laskin, Pieter Abbeel, Aravind Srinivas, and Igor Mordatch. Decision transformer: Reinforcement learning via sequence modeling. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 15084–15097. Curran Associates, Inc., 2021.

[9] Micah Carroll, Orr Paradise, Jessy Lin, Raluca Georgescu, Mingfei Sun, David Bignell, Stephanie Milani, Katja Hofmann, Matthew Hausknecht, Anca Dragan, and Sam Devlin. Uni[MASK]: Unified inference in sequential decision problems. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022.

[10] Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019.