Learning how Humans Learn to Play Board Games with GPT-4IAR

Anonymous submission

Abstract

We present GPT-4IAR, a transformer neural network architecture for modeling and predicting human behavior in the board game four-in-a-row (4IAR). Experiments show that conditioning action predictions on longer histories of previous moves leads to improved accuracy over prior state-of-the-art models, hinting at longer-term strategic biases in human gameplay. Reaction time prediction is also explored, showing promise in capturing meaningful gameplay statistics beyond raw actions. This work ultimately aims to produce a faithful emulator of human cognition to afford detailed investigation into how humans plan and make decisions.

Introduction

Planning and decision making are active areas of research in cognitive science, with great interest in understanding the inner mechanisms underlying how the brain makes decisions in complex, naturalistic scenarios (Hunt et al. 2021). In order to study the cognitive processes and neural structures that dictate how humans decide which action to take in a given scenario, there has been an increasing effort to construct computational models able to describe behavior in different environments and tasks (Collins and Shenhav 2022).

To create accurate cognitive models of planning, *games* have shown to be a great testing ground (Allen et al. 2023). In particular, games offer an environment that feels intuitive and enjoyable for players, while offering a flexible platform that allows researchers to study complex planning through a well-defined set of rules that encode a task. This allows scientists to expand their study to a wider audience that would be attracted to the game, as well as being able to design environments with considerably higher complexity than what has been used previously in psychological studies, while keeping the tasks amenable to scientific analysis (Allen et al. 2023).

In this work, we focus on the game *four-in-a-row* (4IAR), which was previously created to study and develop models around decision-making in a combinatorial game setting (van Opheusden et al. 2017). Models built around this game have studied different aspects of planning; we direct our attention toward the study of expertise and its effects in gameplay (van Opheusden et al. 2023). Our main contribution is GPT-4IAR, a transformer neural network architecture that mimics human behavior in 4IAR. Previous work has used a fully connected neural network to predict the move made by a human player given the current board state (Kuperwajs, Schütt, and Ma 2023). Here, we provide our network with a sequence of previous board states and moves. We show that using the transformer's attention mechanism over previous board states and player's actions improves the network prediction accuracy over the previous state-of-the-art. We also explore the prediction of other human statistics, namely the time it takes someone to make a move ('reaction time'), and show good results toward extending the architecture to do inference on these other gameplay-related statistics.

As a motivation, our ultimate goal is to build a 'perfect emulator' network which is able to mimic the behavior of specific human players, conditioned on a sufficiently long history of previous moves, games, or other summary statistics such as skill level. Such a faithful human emulator could be compared to our best hand-crafted, interpretable cognitive models of learning, planning and decision making (van Opheusden et al. 2023), affording detailed investigations into *how* and *where* our best interpretable models differ from actual human behavior, as a means to push forward our understanding of human cognition. As a byproduct, we would also have models that *play like humans* (as opposed to like an AI), which may have also have practical applications for the gaming industry and more broadly for the field of human-computer interaction.

Background

Task and Dataset

The task under study is a variant of tic-tac-toe called 4-in-a-row (4IAR), where two players aim to connect four tokens on a 4×9 squares board, as proposed by van Opheusden et al. (2017). An example 4IAR board is shown in Figure 1.

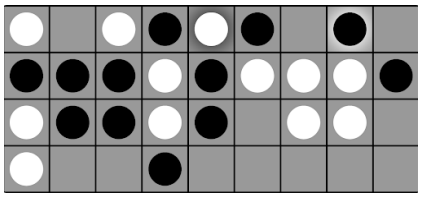

Figure 1: Example board of 4IAR.

This game has approximately 1.2×10^{16} non-terminal states, which provides a level of complexity that far exceeds that of other tasks commonly used in psychology (van Opheusden and Ma 2019). The game is also "simple" enough that building tractable computational models of behavior is still possible (van Opheusden et al. 2017), which can be leveraged to study different aspects of human planning and decision-making, such as expertise (van Opheusden et al. 2023).

For this study, we use a set of approximately 10 million games gathered from a mobile app with a visually enriched version of the game.[1] Users always move first against an AI agent which implements a cognitive model that uses a planning algorithm (van Opheusden et al. 2017; van Opheusden et al. 2023). The AI opponent uses parameters adapted from fits of previously collected human-vs-human games (Kuperwajs, Schütt, and Ma 2023), to yield human-like behavior. The first player (human) is represented by black pieces and the opponent (AI) by white pieces, akin to X's and O's used in tic-tac-toe. Users have a maximum of about ten seconds to make their move, and they end up making, on average, 7.3 moves per game until reaching a termination condition (victory for human, victory for AI, or tie).

Related Work

Transformers in Sequential Decision Making. Transformer-based architectures (Vaswani et al. 2017) have enjoyed great success in tasks with sequential data, with applications in natural language processing, computer vision, and audio processing, to name a few (Lin et al. 2022). By modeling states, actions, and possibly rewards, transformers have also started to enter the domain of reinforcement learning (RL), planning and sequential decision-making (Janner, Li, and Levine 2021; Chen et al. 2021; Carroll et al. 2022). Notably, they have shown great capabilities in imitating human behavior, which may open the door to building computational models in cognitive science and neuroscience for complex tasks (Shafiullah et al. 2022). While we are not specifically building an RL agent, we are interested in replicating and inferring statistics from human behavior in 4IAR through sequences of moves, with the goal of producing a transformer model which we may probe for information to improve the tractable cognitive model to study human planning in the game (van Opheusden et al. 2017).

Neural Networks for Games. Neural network architectures have been used extensively in games, developed to explore strategies in competitive games that have beaten the best human players, with the most prominent example being AlphaZero (Silver et al. 2018; Arulkumaran, Cully, and Togelius 2019). More related to our goals, lately there has been increasing interest in using these models not to find optimal strategies, but to mimic human behavior (McIlroy-Young et al. 2020a,b; Pearce et al. 2023). In a similar fashion to the latter work, we aim to investigate the capabilities of a transformer architecture to predict and emulate human behavior in 4IAR, with particular interest in observing long-horizon dependencies in gameplay (e.g., does conditioning the action on ten – or a hundred – previous moves, possibly even from previous games, yield a better prediction than using only the current board state?). Previous research has been done in using transformers to play chess (Noever, Ciolino, and Kalin 2020) and go (Ciolino, Kalin, and Noever 2020), but they do not focus on the "human" aspect of the game, giving more emphasis to the transformers' capabilities of actually playing the game.

Prior Work on 4-in-a-row. 4IAR has been used mainly to build a tractable computational model of human cognition, developed with hand-crafted, interpretable parameters to explain the decision-making process (van Opheusden et al. 2017). The model has been leveraged to study the interaction of different RL systems (Kuperwajs, van Opheusden, and Ma 2019) and how expertise affects gameplay (van Opheusden et al. 2023). However, in hand-crafting a model, there may be complex patterns that can be missed; hence, it is of interest to have an "oracle" that serves as a perfect emulator of the true cognitive process that we can use to trace such patterns. For this objective, a fully-connected neural network has been previously proposed (Kuperwajs, Schütt, and Ma 2023), which showed a significantly better performance in predicting the next move that a human player would make, compared to the hand-crafted cognitive model. A major limitation of this work is that predictions are based only on the current, single board state. While the current board state is enough to make optimal predictions, it is not enough to predict which move a (specific) human would make. Thus, we seek to go beyond this study and single-moves, using a transformer to study possible longer-horizon dependencies, i.e. if the network can predict the next move better if it has more knowledge from the past history of moves of the human player than just the present board – the history coming from the current game as well as from previously played games. Additionally, we extend the network task to predicting the *reaction time*, i.e., the amount of time that a player takes to make a move, and not just the move itself. For example, this may be of interest to test the idea that players that have more experience will be able to make decisions faster, and to see if there is a discernible pattern in how long they take based on the move or other factors.

Methods

In this section, we describe the implementation details of GPT-4IAR, the transformer architecture we developed to tackle human behavioral mimicry in 4IAR. In this work, we are interested in predicting two things: the *action*, which is the square where the player is going to place their next piece, and the *reaction time* (RT), which is the amount of time that a player takes to take an action, measured in seconds.

Data Representation

Inspired by the Trajectory Transformer (Janner, Li, and Levine 2021), we decided on a simple tokenization approach to represent the boards, actions and RTs as tokens, which we

[1] Details about the mobile app and dataset redacted for anonymity.

can then feed to a standard GPT architecture with little modification. Figure 2 summarizes the process for the tokenization of one *round*, which is composed of one board state and the corresponding action and RT.

We represent each board state $\mathbf{b}$ as a vector of nine entries, one for each column on the board. Each entry corresponds to a base-3 representation of the respective column, as each square on the board has three possible states (empty, black, white), and each column is given an offset to induce an ordering in the vector representation. Then, each action a is represented by a single scalar. Since the board is composed of $4 \times 9 = 36$ squares, there are 36 possible actions, each of which is represented by a single integer value. Finally, due to the nature of tokenization, each RT t must be discretized. To do this, we took the full empirical distribution of RTs from the entire dataset and binned it into twenty quantiles of equal probability, each having 5% of the total probability mass. We can then use these quantiles as boundaries for each of our bins, which we use to decide which token to assign to a given t. Thus, for GPT-4IAR one round comprises of 11 tokens: 9 tokens for the board state, 1 token for the action, and 1 token for the RT.

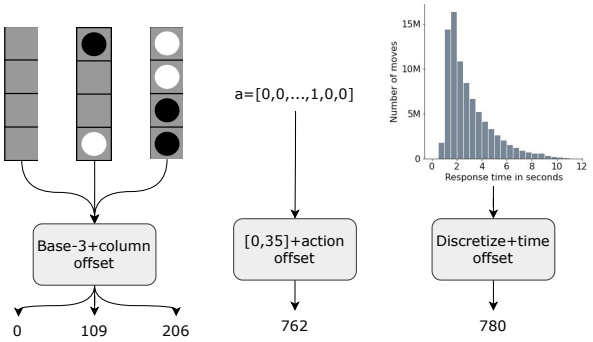

Figure 2: Tokenization scheme of board states, actions, and discretized reaction times (RTs).

We define a *trajectory* τ, following the convention used by the Trajectory and Decision Transformers, as a sequence of rounds:

$$\tau = (\mathbf{b}_1, a_1, t_1, \ldots, \mathbf{b}_T, a_T, t_T)\,,$$

where the indices $i \in [1, T]$ indicate the number of the round played, and T is the current or latest board state. Note that in our representation the trajectory only includes the rounds – board $\mathbf{b}_i$ and actions a_i, t_i – of the human user (black pieces). The next board in the sequence, $\mathbf{b}_{i+1}$, already includes the move of the AI opponent (white pieces) as a response to what the user did in the i-th round.

Network Architecture

A general diagram of the architecture of GPT-4IAR is shown in Figure 3. Essentially, we follow the architecture of GPT-2 (Radford et al. 2019), with the sequence of bespoke tokens we have described in the previous section replacing the string-based tokenization used by text-based large language models.

As a training objective, we use a weighted mean cross-entropy loss, assigning a weight of 1 to the action and RT tokens, and $\frac{1}{9}$ to each of the nine board state tokens. Even though we are not interested in predicting board states per se, our preliminary analyses showed that including board states in the learning objective with a small weight achieves overall better predictive performance than having no weight (and also better than unit weight), possibly by helping the network learn an explicit board representation as well as some opponent modeling.

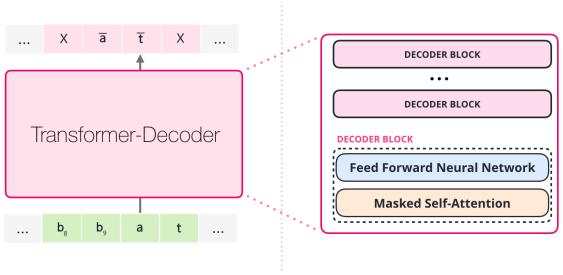

Figure 3: GPT-4IAR architecture. The transformer network is trained on predicting the next token in the sequence (board state, actions and RTs). Figure adapted from "The Illustrated GPT-2" (Alammar 2019).

To train the network, we use a 90/5/5 train/validate/test split on our dataset of 10 million games. We use the AdamW optimizer (Loshchilov and Hutter 2019) to minimize the target loss with parameters $\alpha = 6 \cdot 10^{-4}$, $\beta_1 = 0.9$, $\beta_2 = 0.95$ and $\lambda = 0.1$, which are the default parameters of the open-source implementation we base GPT-4IAR on.

For model assessment, we go through the whole test set to gather the evaluation metrics. Given a board $\mathbf{b}$, the network outputs a probability distribution over all action tokens which is used to compute the cross-entropy loss. We also pick the most likely token as our action prediction used to compute accuracy. Similarly, we input a board and an action $(\mathbf{b}, a)$ to extract a probability distribution over the RT tokens. To measure accuracy, we pick the most likely RT. Accuracy rate may not be the best metric by which to assess RT prediction, since RT is a (discretized) metric continuum. As such, we also evaluate RT prediction through root-mean-square error (RMSE), i.e., distance between our model prediction and the data. For this, we calculate the expected value of the RT token as our prediction and then compute the RMSE with respect to the true RT for each data point, and we aggregate all errors with an average. Losses are calculated separately for action tokens and RT tokens, which are averaged over the test set to give the final reported values.

For all experiments, unless stated otherwise, the network hyperparameters were fixed to the standard GPT-2 values shown in Table 1.

Experiments

In this section we report our preliminary results with GPT-4IAR. While thorough statistical testing and additional experiments are needed to draw more definite conclusions, several trends can already be observed in our experiments.

Hyperparameter	Value
Embbedding dimensionality	768
Layers	12
Attention Heads	12

Table 1: Fixed hyperparameters used for training.

Training Networks with Different Context Lengths

In order to evaluate the performance of GPT-4IAR at predicting human behavior when more past information is potentially available both during training and at test time, we trained three networks that differed in their *context length* (also known as context window) – the maximum token sequence that the network can process –, all else being equal. We trained three networks with context lengths of 256, 512, and 1024 tokens, respectively. The loss curves from training are shown in Figure 4.

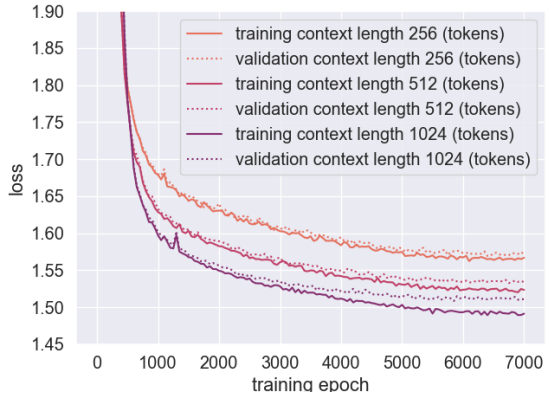

Figure 4: Observed training and validation loss for GPT-4IAR networks with different context lengths.

The curves show an asymptotic reduction of both training and validation loss as a function of context length (lowest loss achieved by the different colored lines), suggesting that the network is able to extract more information from observing further into the past to improve predictions. Moreover, it seems likely that further extending the context length would still improve performance. However, in practice there are well-known computational limitations to implementations with longer context windows due to the quadratic scaling of the standard attention mechanism as a function of context length (Vaswani et al. 2017).

Assessing Performance Under Different Sequence Lengths

Now that we have trained networks with different context lengths, we can evaluate the role of *sequence length* on performance at inference time, by systematically varying the length of the sequence of past rounds the network can use to make predictions. Clearly, the maximum number of rounds each network can process is limited by its context length. Remember that a round is 11 tokens long, so our networks can store in context from up to 23 rounds for our smallest network to up to 93 rounds for our largest network. Considering that a game is on average 7.3 rounds in our dataset, even our smallest network can store in context a few past games.

Moreover, we provide comparisons with the previous state-of-the-art prediction results (Kuperwajs, Schütt, and Ma 2023), particularly in terms of accuracy on next-move-prediction on the test set, and some qualitative assessments on the output of the network on single boards.

Action Prediction. The accuracy of action prediction as a function of the size of a trajectory of past game states, actions and RTs received as context is shown in Figure 5. We observe a substantial, positive effect on prediction accuracy of increasing the sequence length. In particular, we observe an improvement of up to around $6-7\%$ (from about 41% to 48% accuracy) when we include more past rounds into the context compared to only having one round (the current board). These results ostensibly indicate that the transformer is able to better predict behavior based on long-term dependencies in decisions. Notably, performance does not seem to be plateauing, suggesting that networks could be able to exploit even longer temporal correlations.

Finally, while all GPT-4IAR models exhibit similar performance when evaluated on sequences of the same length, there is perhaps a small advantage in using larger networks trained on longer contexts. This result, if confirmed, would suggest that networks trained on longer contexts are better even when limited to short sequences. More analyses are needed to assess statistical significance of this finding.

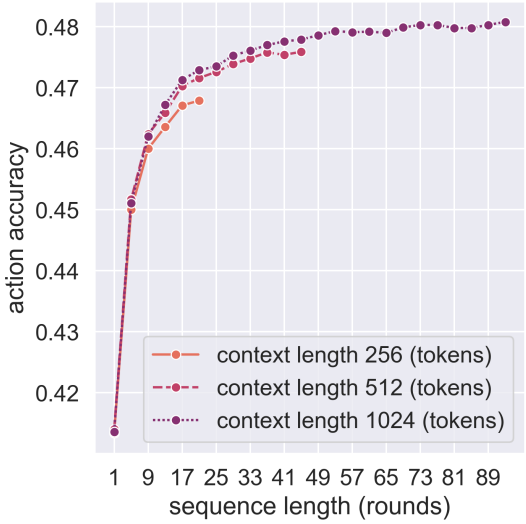

Figure 5: Action prediction accuracy of GPT-4IAR models with different context length as a function of provided sequence length (one round = 11 tokens).

We compare our best performing GPT-4IAR model against the previous state-of-the-art fully connected network model by Kuperwajs, Schütt, and Ma (2023) in Table 2.

Some examples of the actions predicted by GPT-4IAR are

Prediction	Metric	Fully Connected	GPT-4IAR
Actions	Accuracy	41.71 %	48.08 %
	Loss	1.866	1.504
RTs	Accuracy	—	14.69 %
	Loss	—	1.508
	RMSE	—	5.16 bins

Table 2: Comparison between the fully connected model and GPT-4IAR at predicting actions and RTs, using different metrics: accuracy, cross-entropy loss, and root mean squared error (RMSE). Only GPT-4IAR predicts RTs.

shown in Figure 6. Qualitatively, observing Figure 6(a), we can see that the network is able to capture lapses in human gameplay. The optimal move would be to block white from connecting four on a diagonal, but we can see a low, but non-zero probability of making moves that would try to develop a win condition for black. In the board shown in Figure 6(b) we can also see that the network is able to capture uncertainty on "harder" boards too, with diffuse probabilities across the board, e.g. on the second row, seventh and ninth column, to make a decision.

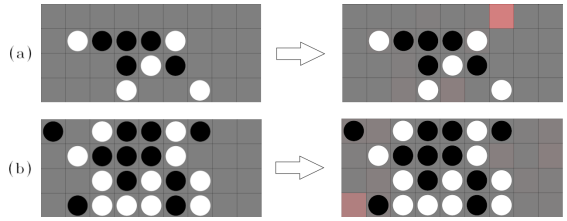

Figure 6: Example distributions of predicted moves (right) given two distinct input boards (left). Intensity of red in each square indicates the action probability assigned by the GPT-4IAR network. See text for discussion.

Reaction Time Prediction. We also evaluate the performance of the network at predicting the reaction time of a player. First, we measure prediction accuracy by choosing the most likely RT token given a trajectory of past boards, moves and RTs. RT prediction accuracy improves with the length of the provided sequence, as shown in Figure 7, reaching a maximum of approximately 14.69%.

Since RT is a continuous variable, accuracy may not necessarily be the best measure of performance for prediction, e.g. it may still be acceptable if we predict one bin up or down from the most likely value. Hence, we also study the RMSE of RT prediction, measured in terms of bin distance, shown in Figure 8. Similarly to the accuracy results, RMSE improves as a function of sequence length. For reference, the RMSE of the RT data with respect to a constant prediction is 6.66 bins.

Discussion

In this paper, we introduced GPT-4IAR, a transformer architecture to predict human behavior in a board game setting. We showed that there can be significant information

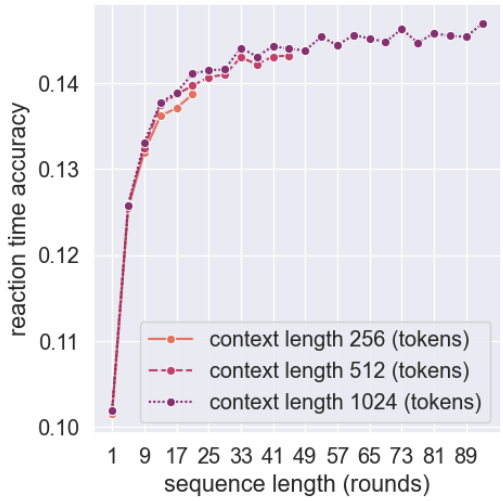

Figure 7: Reaction time prediction accuracy of GPT-4IAR models with different context length as a function of provided sequence length (one round = 11 tokens).

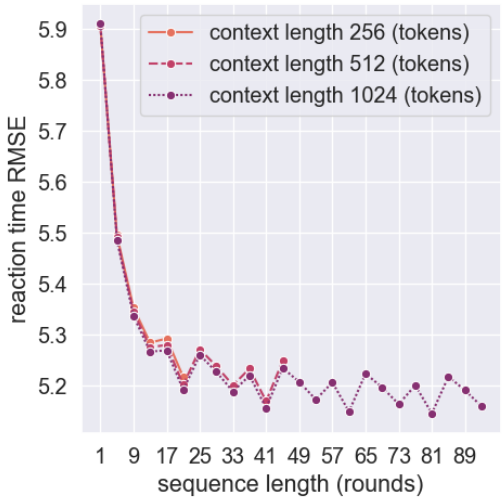

Figure 8: Reaction time root mean squared error (RMSE) as a function of provided sequence length, measured in bins.

gain in predicting the decisions of a human agent by taking their past behavior into account, yielding better performance than previous state-of-the-art fully connected networks that only took a single board into account to make a decision. This hints at significant correlations on a longer horizon of moves, meaning that a personal bias in strategy or other latent variables can influence gameplay over many rounds. An obvious example of an underlying latent variable influencing gameplay that may be inferred over multiple games – and might not be immediate by only observing a short sequence – is the player's skill or expertise.

We also showed that there may be a correlation between

the time a player takes to make a move and their long-term strategies. We show this quantitatively through prediction accuracy and RMSE of (discretized) reaction times. This result could also hint at individual features, such as skill, that the network may be able to identify through long trajectories of game states and actions.

Further immediate work includes more thorough statistical testing to consolidate our conclusions, performing ablation tests with hyperparameters other than maximum context length, and exploring other representations for the data to see if there can be improvements over the "naive" token representation of the board.

In the future, we plan to explore the inference of other statistics, such as Elo score (a measure of the player's skill), from which we could further extract information on what affects human decisions. Going a step further, we could also *condition* the predicted gameplay by feeding the network with relevant statistics, such as the player's Elo score. In other words, we could train a single network that would be able to play like humans with different levels of expertise. Notably, previous efforts such as Project Maia achieved this feature by training distinct models for different Elo tiers (McIlroy-Young et al. 2020b).

Finally, the framework we presented and our results with this specific game can likely be extended to other combinatorial games and tasks, such as Go and Chess, as the architecture just depends on the tokenization. Although previous work has addressed the use of transformers to play Chess and Go (Noever, Ciolino, and Kalin 2020; Ciolino, Kalin, and Noever 2020), this was done through fine-tuning a text-based GPT to a restricted vocabulary corresponding to algebraic notation on each game. It would be interesting to explore potential differences in performance between these different training approaches.

Acknowledgements

For the anonymous submission we do not name our acknowledgements.

Code References

We used the NumPy (Harris et al. 2020) and PyTorch (Paszke et al. 2019) open-source libraries for this work. The architecture used is heavily based on Andrej Karpathy's open source implementation of GPT-2, called nanoGPT (Karpathy 2023).

Code Availability

For the anonymous submission we do not link the repository with our work.

Dedicated Section

In this work, we presented GPT-4IAR, a transformer neural network architecture that aims to emulate human behavior in the cognitive task of playing the four-in-a-row (4IAR) board game. Our work sits exactly at the intersection of the fields of cognitive science, machine learning, and human-machine interaction or collaborative AI.

From the cognitive science perspective, our 'oracle' model can be used by cognitive scientists as a *virtual laboratory* to test hypotheses and improve our best hand-crafted cognitive models of human learning, planning and decision making within this board game setting, by having access to 'virtual humans' that can be tested in unlimited scenarios and counterfactual conditions. Clearly, results obtained *in silico* with emulated humans would need to be reproduced with real humans, but still we envision this could substantially speed up research into human cognition by systematically exploring non-trivial shortcomings in our state-of-the-art cognitive models of game playing (van Opheusden et al. 2023).

In terms of machine learning, the current work is based on established transformer neural network architectures (Vaswani et al. 2017; Radford et al. 2019). Still, once our approach is scaled to more complex settings and higher fidelity of human emulation we envision several challenging problems will arise which will require the development of advanced machine learning solutions. To name one, we will potentially have to deal with scaling our method to very long context windows, an open research area within modern large language models (Press, Smith, and Lewis 2022; Su et al. 2024).

Finally, our work has immediate implications for human-machine interaction and collaborative AI. Our 'human emulators' can be used as better opponents for other human players – playing like humans and not like AIs. While work with similar goals exists such as Project Maia (McIlroy-Young et al. 2020b), our model will extend beyond that by potentially affording specific emulation of a variety of different game styles within the same model. More broadly, our approach is an example of building an AI system that behaves like a human (McIlroy-Young et al. 2020a,b; Pearce et al. 2023), which can be useful both for building collaborative AI models (using emulated humans as a proxy during training) as well as a technique for making the trained AI models themselves behave more like humans.

References

Alammar, J. 2019. The Illustrated GPT-2 (Visualizing Transformer Language Models). https://jalammar.github.io/illustrated-gpt2.

Allen, K. R.; Brändle, F.; Botvinick, M.; Fan, J.; Gershman, S. J.; gopnik, a.; Griffiths, T. L.; Hartshorne, J. K.; Hauser, T. U.; Ho, M. K.; and et al. 2023. Using Games to Understand the Mind.

Arulkumaran, K.; Cully, A.; and Togelius, J. 2019. AlphaStar: An Evolutionary Computation Perspective. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, GECCO '19, 314–315. New York, NY, USA: Association for Computing Machinery. ISBN 9781450367486.

Carroll, M.; Paradise, O.; Lin, J.; Georgescu, R.; Sun, M.; Bignell, D.; Milani, S.; Hofmann, K.; Hausknecht, M.; Dragan, A.; and Devlin, S. 2022. Uni[MASK]: Unified Inference in Sequential Decision Problems. In Oh, A. H.; Agar-

wal, A.; Belgrave, D.; and Cho, K., eds., *Advances in Neural Information Processing Systems*.

Chen, L.; Lu, K.; Rajeswaran, A.; Lee, K.; Grover, A.; Laskin, M.; Abbeel, P.; Srinivas, A.; and Mordatch, I. 2021. Decision Transformer: Reinforcement Learning via Sequence Modeling. In Ranzato, M.; Beygelzimer, A.; Dauphin, Y.; Liang, P.; and Vaughan, J. W., eds., *Advances in Neural Information Processing Systems*, volume 34, 15084–15097. Curran Associates, Inc.

Ciolino, M.; Kalin, J.; and Noever, D. 2020. The Go Transformer: Natural Language Modeling for Game Play. In *2020 Third International Conference on Artificial Intelligence for Industries (AI4I)*, 23–26.

Collins, A. G. E.; and Shenhav, A. 2022. Advances in modeling learning and decision-making in neuroscience. *Neuropsychopharmacology*, 47(1): 104–118.

Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; van Kerkwijk, M. H.; Brett, M.; Haldane, A.; del Río, J. F.; Wiebe, M.; Peterson, P.; Gérard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; and Oliphant, T. E. 2020. Array programming with NumPy. *Nature*, 585(7825): 357–362.

Hunt, L. T.; Daw, N. D.; Kaanders, P.; MacIver, M. A.; Mugan, U.; Procyk, E.; Redish, A. D.; Russo, E.; Scholl, J.; Stachenfeld, K.; Wilson, C. R. E.; and Kolling, N. 2021. Formalizing planning and information search in naturalistic decision-making. *Nature Neuroscience*, 24(8): 1051–1064.

Janner, M.; Li, Q.; and Levine, S. 2021. Offline Reinforcement Learning as One Big Sequence Modeling Problem. In *Advances in Neural Information Processing Systems*.

Karpathy, A. 2023. nanoGPT. https://github.com/karpathy/nanoGPT.

Kuperwajs, I.; Schütt, H. H.; and Ma, W. J. 2023. Using deep neural networks as a guide for modeling human planning. *Scientific Reports*, 13(1): 20269.

Kuperwajs, I.; van Opheusden, B.; and Ma, W. J. 2019. Prospective planning and retrospective learning in a largescale combinatorial game. In *2019 conference on cognitive computational neuroscience*, 13–16.

Lin, T.; Wang, Y.; Liu, X.; and Qiu, X. 2022. A survey of transformers. *AI Open*, 3: 111–132.

Loshchilov, I.; and Hutter, F. 2019. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations*.

McIlroy-Young, R.; Sen, S.; Kleinberg, J.; and Anderson, A. 2020a. Aligning Superhuman AI with Human Behavior: Chess as a Model System. In *26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*.

McIlroy-Young, R.; Wang, R.; Sen, S.; Kleinberg, J.; and Anderson, A. 2020b. Learning Personalized Models of Human Behavior in Chess.

Noever, D.; Ciolino, M.; and Kalin, J. 2020. The Chess Transformer: Mastering Play using Generative Language Models. arXiv:2008.04057.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; Desmaison, A.; Kopf, A.; Yang, E.; DeVito, Z.; Raison, M.; Tejani, A.; Chilamkurthy, S.; Steiner, B.; Fang, L.; Bai, J.; and Chintala, S. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*, 8024–8035. Curran Associates, Inc.

Pearce, T.; Rashid, T.; Kanervisto, A.; Bignell, D.; Sun, M.; Georgescu, R.; Macua, S. V.; Tan, S. Z.; Momennejad, I.; Hofmann, K.; and Devlin, S. 2023. Imitating Human Behaviour with Diffusion Models. In *The Eleventh International Conference on Learning Representations*.

Press, O.; Smith, N.; and Lewis, M. 2022. Train Short, Test Long: Attention with Linear Biases Enables Input Length Extrapolation. In *International Conference on Learning Representations*.

Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; and Sutskever, I. 2019. Language Models are Unsupervised Multitask Learners.

Shafiullah, N. M. M.; Cui, Z. J.; Altanzaya, A.; and Pinto, L. 2022. Behavior Transformers: Cloning k modes with one stone. In *Thirty-Sixth Conference on Neural Information Processing Systems*.

Silver, D.; Hubert, T.; Schrittwieser, J.; Antonoglou, I.; Lai, M.; Guez, A.; Lanctot, M.; Sifre, L.; Kumaran, D.; Graepel, T.; Lillicrap, T.; Simonyan, K.; and Hassabis, D. 2018. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, 362(6419): 1140–1144.

Su, J.; Ahmed, M.; Lu, Y.; Pan, S.; Bo, W.; and Liu, Y. 2024. RoFormer: Enhanced transformer with Rotary Position Embedding. *Neurocomputing*, 568: 127063.

van Opheusden, B.; Galbiati, G.; Bnaya, Z.; Li, Y.; and Ma, W. 2017. 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, 1254–1259. The Cognitive Science Society.

van Opheusden, B.; Kuperwajs, I.; Galbiati, G.; Bnaya, Z.; Li, Y.; and Ma, W. J. 2023. Expertise increases planning depth in human gameplay. *Nature*, 618(7967): 1000–1005.

van Opheusden, B.; and Ma, W. J. 2019. Tasks for aligning human and machine planning. *Current Opinion in Behavioral Sciences*, 29: 127–133. Artificial Intelligence.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L. u.; and Polosukhin, I. 2017. Attention is All you Need. In Guyon, I.; Luxburg, U. V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; and Garnett, R., eds., *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.