The Reinforcement Learning (RL) field has seen huge advancement in recent years. such as agent 57 [2] which could play all 57 atari games and the generalist agent[6] which could perform a wide range of tasks. Despite the improvement, they are still brittle in terms of generalization, and efficiency that prevents massive real-world application, they are unable to maintain similar performance when moving to a similar unseen setting where humans can adapt to that quickly and still spend a lot of time on the relative hard-to-explore game yet seems easy in human's perspective. These problems motivate me to explore **how to leverage the knowledge that we learned from ourselves (neuropsychology) and causality to improve RL's efficiency and generalizability**. I'm fortunate to have conducted some initial research in these areas under the guidance of Prof. Julian Togelius alongside other wonderful collaborators, and I am passionate to continue studying these problems in my Ph.D.

Neuropsychology-inspired RL for efficiency and generalizability. Neuropsychology has lots of connections to reinforcement learning. One important aspect I want to discover is how agents could learn from humans to make decisions when no external reward is provided which is rather common in our life. Many current RL algorithms rely on the hardcoded reward function. The problem with it is it only gives rewards for some specific states, and with some predefined number, in contrast, humans will consistently motivate themselves internally when making sequential decisions, For instance, when learning to play a game after watching others' demonstrations. Humans will prefer to take some actions when they could reach to same or similar state that they saw from others' good demonstrations, and avoid them if the demonstrations are bad even though the game does not provide any reward for these transitions. They will also have a preference on the action choice based on the goal even if the game does not return any reward for these state-action pairs. Motivated by how humans learn from others. Draw inspiration from [1] on my first RL project: finding a policy that could match the tree search agent performance on grid-based games, **I designed an algorithm to train an agent by providing extra reward when the current state is close to learned play trace embedding**. The trained agent was able to win some levels which proves the feasibility of improving performance by imitating human demonstrations. From that experiences, I learned how RL works and am excited to **explore ways of improving agent efficiency by designing a neuropsychology-inspired reward function**. Among different type of intrinsic motivations, curiosity might be one of the most important factor to human society's advancement. RL community already adopts it to improve the performance on hard-exploration games, and it solved some of the problems that hardcoded reward function is facing, but is prone to be attracted by unrelated examples, for instance, in random network distillation [3], when they put a random noisy tv on the wall, it completely attracts the agent which is totally irrelevant to the game's goal. Observing that the agent keeps getting high rewards for watching noisy TV, **I proposed to use the difference of intrinsic reward from 2 consecutive steps to construct a new reward**. The preliminary results show similar performance and I would like to **further probe the model performance under the random noisy tv setting and to explore ways to construct a better curiosity reward for these situations, especially how to incorporate it with the goal so that agent will not be attracted by irrelevant objects**.

Visual information is another crucial factor in human's decision-making process. Humans usually attend to one or several area in the scene for making decisions. Many RL agents also involve visual input. But most of the RL algorithms are using standard 3 layers conv-net from nature atari paper [5] which provides equal weights to all regions of the input image at the beginning. And that results in the failure or slow-to-learn in some pretty simple games. To leverage the attention mechanism from humans. **I proposed an algorithm that forces the agent to only focus to local view for grid-based games** [9] as in the grid-based game, local information plays a crucial role in winning the game. By manually cropping the input frame to a local ego-centric view, we force the agent to attend to important areas. The result proves that the local view provides more information for winning the game than a global view without any attention. But it comes with a problem, even though the models be forced into an ego-centric view, we need to access the game engine to retrieve an ego-centric view. Besides, the ego-centric view might lose other information when applied to different settings such as RTS where key information might be spread out on the screen but not limited into one small area. With that question, we **proposed a self-supervised representation learning model that is weakly supervised by pseudo-spatial human attention labels generated by a teacher model trained on human saliency data** [8]. The model learns to attend areas where humans are likely to attend

while also boosts the classification and image retrieval performance. I'm interested to **utilize such model to further improve RL model's efficiency and generalizability**.

Causal representation learning in RL. Beyond neuropsychology-inspired RL, I'm also getting more interest in casual representation learning in RL as I found out that even though agents that combine an image representation learning model show some generalization performance on the unseen level within the same game, it will still completely fail on a similar setting game, Unlike that, humans could quickly adapt to new settings as they share a similar casual graph, such as driving a car in different games [7]. Currently, causality has not gained a lot of attention in RL community yet. But I think it could play a crucial role in advancing the generalizability of RL model. Taking the CuRL [4] as an example. Its visual backbone computes the representation purely from images. But humans would also include other information such as action, environment and context into representation when making a decision. I'm a complete newcomer in this field but I can see the connection between casual representation learning and RL that can be used to improve the generalizability of RL agents and would like to explore this area.

*References

[1] Yusuf Aytar, Tobias Pfaff, David Budden, Thomas Paine, Ziyu Wang, and Nando De Freitas. Playing hard exploration games by watching youtube. *Advances in neural information processing systems*, 31, 2018.

[2] Adrià Puigdomènech Badia, Bilal Piot, Steven Kapturowski, Pablo Sprechmann, Alex Vitvitskyi, Zhaohan Daniel Guo, and Charles Blundell. Agent57: Outperforming the atari human benchmark. In *International Conference on Machine Learning*, pages 507–517. PMLR, 2020.

[3] Yuri Burda, Harrison Edwards, Amos Storkey, and Oleg Klimov. Exploration by random network distillation. *arXiv preprint arXiv:1810.12894*, 2018.

[4] Michael Laskin, Aravind Srinivas, and Pieter Abbeel. Curl: Contrastive unsupervised representations for reinforcement learning. In *International Conference on Machine Learning*, pages 5639–5650. PMLR, 2020.

[5] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533, 2015.

[6] Scott Reed, Konrad Zolna, Emilio Parisotto, Sergio Gomez Colmenarejo, Alexander Novikov, Gabriel Barth-Maron, Mai Gimenez, Yury Sulsky, Jackie Kay, Jost Tobias Springenberg, et al. A generalist agent. *arXiv preprint arXiv:2205.06175*, 2022.

[7] Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. Toward causal representation learning. *Proceedings of the IEEE*, 109(5):612–634, 2021.

[8] Yushi Yao, Chang Ye, Junfeng He, and Gamaleldin Fathy Elsayed. Teacher-generated pseudo human spatial-attention labels boost contrastive learning models. In *SVRHM 2022 Workshop@ NeurIPS*.

[9] Chang Ye, Ahmed Khalifa, Philip Bontrager, and Julian Togelius. Rotation, translation, and cropping for zero-shot generalization. In *2020 IEEE Conference on Games (CoG)*, pages 57–64. IEEE, 2020.