

Training human-like bots with Imitation Learning based on provenance data

Lauro Víctor Ramos Cavadas
Instituto de Computação
Universidade Federal Fluminense
Niteroi, Brazil
laurovrc@id.uff.br

Esteban Clua
Instituto de Computação
Universidade Federal Fluminense
Niteroi, Brazil
esteban@ic.uff.br

Troy Costa Kohwalter
Instituto de Computação
Universidade Federal Fluminense
Niteroi, Brazil
troy@ic.uff.br

Sidney Araujo Melo
Instituto de Computação
Universidade Federal Fluminense
Niteroi, Brazil
sidneymelo@id.uff.br

Abstract—Believable NonPlayer Characters in video games are one of the most challenging problems in the game industry over the last years. Players demand to expect to perceive the NPCs as other human-based player. Modeling NPC behavior manually is not always a good choice, mainly due to the number of NPCs a game can have and the difficulty of modeling a large number of actions that they can take. Our main goal is create a believable NPC acting like a real player. This work proposes an approach to training an NPC using Imitation Learning so that it is as similar as possible to a human player. Through this strategy, NPCs are trained from various types of players, avoiding predefined behaviors. Our proposal trains agents with the use of provenance data sets, tackling cause-effects data mining possibilities, and use Generative Adversarial Imitation Learning framework to take actions similar to what a player would take. The model proposed was create to be generic and applicable to various games. We validate our presented model with the DodgeBall environment inside Unity ML-Agents Toolkit for Unity Engine. Some players was asked to play against our agent and they validated the believability of our trained NPCs.

Index Terms—NPC, Imitation Learning, Provenance, Games

I. INTRODUCTION

Non-Player Characters (NPCs) are autonomous agents in video games, which interact with the game environment and the player. They are therefore an essential element of gameplay and for the player's immersive feeling. Traditionally, NPC behavior creation methods are manual, requiring substantial effort and mastery of several areas, such as scripting and testing, in order for it to work properly. In addition, a populous world that aims to provide a good immersion for the player has a huge amount of NPCs and for a player, each NPC has to be more believable, engaging, and human-like as possible.

The imitation of human playing style has been gaining relevance over the last few years. The goal of these agents is to deceive real players and be perceived just as other human player. In recent years, imitation learning has been investigated as a way to efficiently and intuitively program autonomous behavior [1]-[6]. Numerous imitation learning methods have

been developed and imitation learning has become a strong field of research.

The purpose of Imitation Learning (IL) is to efficiently learn the desired behavior by imitating an expert's behavior. So this work propose a model for training a believable NPC using IL that does not require an expert to teach the NPC. Thus, the NPC will not behave with perfect mastery as we will not use an expert to train it. Instead we will use provenance data gathered from gameplays of multiple matches from many different players. Provenance has been used for recording documented history of an object's life cycle and is generally used in the context of art, digital data, and science [7]. Its first usages for game telemetry and analytics are described on [8].

Ho and Ermon [9] proposed a framework called Generative Adversarial Imitation Learning (GAIL) that explores randomly to determine which actions bring a policy's occupancy measure closer to the trainer's, rather than methods that interact with him. The main contribution of this work is that our model use data collected from real players gathered with provenance to train believable NPCs using GAIL and this model can be applied to any game.

We choose the game engine Unity3D¹ to train and validate an NPC, using the DodgeBall² game environment present in open source Unity ML-Agents Toolkit³. Fig. 1 shows a picture of the Dodgeball game, showing a player's avatar and the environment with walls and obstacles.

Dodgeball game, which is a team-based game, already has been used for training NPCs through MultiAgent POsthumous Credit Assignment (MA-POCA) [10]. Dodgeball allows the player to adopt different strategies such as being more aggressive or shooting balls from a long distance. An important point in our model is to reward the agent when it take actions

¹<https://unity.com>. Last accessed: 10 Mar 2022

²<https://blog.unity.com/technology/ml-agents-plays-dodgeball>. Last accessed: 15 Apr 2022.

³<https://github.com/Unity-Technologies/ml-agents>. Last accessed: 10 Mar 2022.



Fig. 1. DodgeBall game.

that were performed by a player and were recorded in the provenance, regardless of whether the actions had good or bad results. This will make the agent not perfect and make the same mistakes as a real player.

The remaining of the paper is organized as follows: The second section presents related work and the third section presents our proposed model. The fourth section bring the early results and the last section concludes this work, pointing out future works.

II. RELATED WORK

Machine Learning (ML) techniques can be used to automate the process of learning how to play a video game either progressively using players' game traces as input, through direct imitation approaches, or using some form of optimization technique such as Evolutionary Computation or Reinforcement Learning to develop a fitness function that, for instance, "measures" the human likeness of an agent's playing style [11].

Previous works addressed the problem of imitating human players using different supervised learning approaches. Neuroevolution [12] was used in Ms. Pac-Man and in other game domains [13], and more recently case-based reasoning [14] has been proposed due to its capacity for imitating spatially-aware autonomous agents in a real-time setting [15]. None of these works are based on real players gameplays logs.

Kohwalter et al. [8] proposed a novel approach named PinGU¹ for capturing and storing provenance data from a game session based on the Provenance in Games conceptual framework. The wealth of provenance data collected during a game session is fundamental for understanding the mistakes made as well as reproducing the same results at a later moment. Causal relationships between game elements are mapped as edges connecting their respective nodes, resulting in a game provenance graph. Causality indicates a relationship between two events, where the former event affects the later. The provenance approach capture causal relationships explicitly defined by the game developer. Each edge captured through the provenance approach represents a type of relationship (that can also be causal) between game objects' actions and/or

¹<https://github.com/gems-uff/ping>

TABLE I
FEATURES FROM RELATED WORK AND OUR WORK.

	Use IL	Use log?	Type of log	Focus on...
Kohwalter et al.	No	Yes	Provenance	Provenance
Karpov et al.	Yes	Yes	Raw data	Controllers
Cruz and Uresti	No	Yes	Raw data	Agents
Pelling and Gardner	Yes	Yes	Raw data	Agents
Miranda et al.	No	Yes	Raw data	Agents
Cavadas et al.	Yes	Yes	Provenance	Agents

states. The most important advantages of provenance graphs are the modeling of causal relationships, which structures the provenance elements into a graph, and its richness of detail.

Karpov et al. [16] presented a component of the UT² bot, inspired by the idea of direct imitation of human behavior. The controller draws upon a previously collected database of recorded human games, which is indexed and stored for efficient retrieval. The controller works by quickly retrieving relevant traces of human behavior, translating them into the action space of the bot and executing the resulting actions. The main difference between this work and our proposal is that the controller search for an previous recorded situation. In the case this is not found, the agent will stuck in the same position. Our model uses GAIL framework in order to prevent this situation.

Cruz and Uresti [17] stated two main challenges on their work: The first challenge was exploring domains with high-dimensional state-action spaces, while satisfying constraints imposed by traits that characterize human-like behavior. To approach this problem, the proposed framework learns the model of a game by observing how humans play that game. The second challenge found was generating varied behaviors, which also adapt to the opponent's playing style. The main difference compared to our model is the learning method. While in the framework is used RL in real-time to obtain player's actions, in our model we recreate games from provenance files and train the agent with IL.

Pelling and Gardner [18] discussed two designs for imitation-learning bots. Both were based on support vector techniques and include a novel probabilistic model for in-combat jumping. One bot also includes a application of probability estimation trees to in combat movement. Both designs appeared viable when tested under laboratory conditions and in the competition format of the 2009 2K BotPrize. Table I compares some features from related works and from our work.

III. PROVENANCE BASED IMITATION LEARNING MODEL

In this paper, we propose an Imitation Learning approach, based on provenance data, for training NPCs with believable behaviors instead of manually programming them. We differ from all existing approaches since our model uses provenance data instead of a regular log and the rewards given to the brain during training is not based on give rewards/penalties when the NPC act correctly/incorrectly, instead we give rewards when

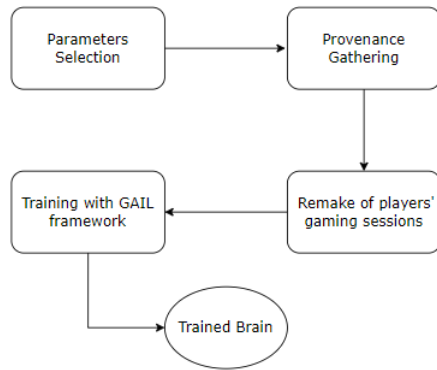


Fig. 2. Presented model.

he does the actions provided by the provenance from players game sessions.

The model proposed was created to generate NPCs that can be applied to a variety of game types where the NPC and player has the same actions, like in a racing game or a First Person Shooter game. Note that the model described in this section doesn't need specific configuration of the game, we only use DodgeBall game as a case of use.

Our solution is composed of four stages: (1) choose the parameters required for recreating player actions, (2) Gathering player provenance data sets, (3) Remake and training of players' gaming sessions based on provenance data, and (4) training with GAIL framework. As a result, we produce a brain generated with MI-Agents and provenance data that can be used on the NPC agents, mimicking a human-like behavior. The model is illustrated in Fig. 2.

Before executing the described steps, it is necessary to prepare the game environment, in our case the DodgeBall game. This is made through a special gameplay setup, where the NPC's behaviors are substituted by the main player. The game has 4 players on the first team and 4 players on the second team. We change the number of players leaving only 1 player at each team. We also change the behavior that control the victory/defeat conditions so that only one player may give the victory to the team.

A. Parameters selection

The first stage of the method consists on verifying and choosing which gameplay's information and which inputs will be required to be captured by the provenance. It is necessary to save all inputs referring to the actions that the bot will be able to do. The parameters are mapped as the actions that our NPC can do. As an example, one of the inputs that will be saved is the input referring to the action of throwing the ball at an opponent. However, to play a ball the player must be carrying at least one ball. Any information that is connected to an action will also be collected by provenance. In order to capture provenance data from a game, we use PinGU.

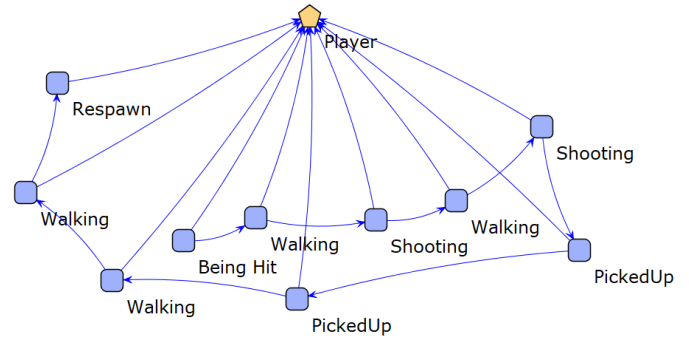


Fig. 3. Snippet of some activities that were saved in the provenance file.

B. Gathering player session data via provenance

In this stage we include at the player's action all the gathered provenance data, which is represented in a graph containing vertices representing player actions during the game session, edges that are the relationships between actions, agents and events, and the game state information during each moment an action was recorded. We follow five stages to get data using PinGU: we first create a game object in the scene that acts as a centralizing server for the provenance information. This game object will have two attached classes: ProvenanceController and InfluenceController. Both classes are used to manage all provenance information and graph generation, thus only one instance of each are necessary per game scene. The second stage is to attach the ExtractProvenance class in the player's agent in the game and link it to the object created in the first step. This class is responsible for creating all the provenance nodes for the game entity and then passing these nodes to the ProvenanceController in order to be inserted at the graph. The third stage consists on identifying the actions and their interactions with other actions that we want to map. For instance, in our implementation, we identify walk, rotate, throw the ball, being hit, pick a ball from the ground and win/lose. The fourth stage is creating the domain-specific provenance tracking functions and attaching it in the player's agent. The object should have a provenance function for each possible action that can be performed and that we are interested for tracking. Finally, it is necessary to add a provenance export function to an event so it can save the current provenance graph to an external xml file. A small snippet of the provenance file containing some activities collected during the game can be viewed in the Fig. 3.

After finishing this configuration, it is time to collect the data of the players who will play the game. For validation purposes, a player is in control of an NPC agent and will play against an AI-controlled agent that will be on the opponent team. The player will play several matches and each match will generate data that will be saved in the provenance file. In the end, we will have several game sessions from a single player. It is possible to collect and merge data from a large set of different players and users.

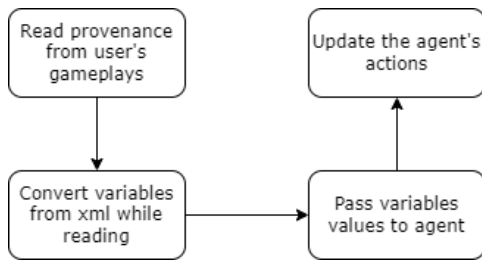


Fig. 4. Steps required for remaking actions saved on provenance.

C. Recreating players' gaming sessions based on provenance for training

We include variables responsible for each action of the character executed by the human trainer in the agent's control behavior. For instance, in our implemented scenario they are: move forward, move back, move left and move right, and throw the ball.

It is important to note that when the player makes an input, the agent's control behavior checks which input was triggered and, according to it, fills the variable created to represent each action. At the end of each frame, the values contained in these variables are passed to a function, which checks the values of the variables and performs the correct actions. For example, when the player presses the key responsible for making the character move forward, the value of the variable responsible for moving the character forward receives a positive value. At the end of the frame, this variable is checked, and as it has a positive value, the character moves forward.

We start this stage by reading the provenance file. We must read each entry from the XML file, node by node, and interpret the data. A new node will be read and recorded at each frame. At this moment the XML serialization is used for converting an object's public properties and fields to a serial format (in this case, XML) for storage or data transport. A Deserialization re-creates the object in its original state from the XML output. It is possible to understand the serialization as a way of saving the state of an object into a stream or buffer.

The interpretation is executed after reading the script. At each node from the source file we obtain the values referring to each input. This data is converted to inputs, that are then converted from a string format within the XML to a float format, related to the variables responsible for the character's actions. Finally, at each frame, the value of these variables is verified and the function responsible for executing the character's actions is executed. This process is represented in Fig. 4.

D. Training with the GAIL framework

At this stage, the game is ready to recreate the contents of a provenance file. Before starting the game and recreating the actions coming from the provenance, it is necessary to configure how the agent will be trained. For this, we created a neural network configuration file for training, including the configuration for using the GAIL framework. Unity ML-

```

network_settings:
  normalize: false
  hidden_units: 512
  num_layers: 3
  vis_encode_type: simple
reward_signals:
  gail:
    gamma: 0.99
    strength: 1.0
  network_settings:
    normalize: true
    hidden_units: 128
    num_layers: 2
    vis_encode_type: simple
  learning_rate: 0.0003
  use_actions: false
  use_vail: false
  demo_path: Demos/DemoFinal.demo
  keep_checkpoints: 40
  
```

Fig. 5. Snapshot from training configuration file.

Agents Toolkit already has GAIL implemented so we only configure the network parameters, like strength and learning rate. 5 shows a snapshot from the training configuration file, including GAIL configuration. This step can use several game sessions recorded from the same player or from multiple users. In case of using multiple users, the process starts with the first player's data and the training will be incremented with all the others player data.

When the game starts we begin to train and we have to give rewards/penalties to the brain so that it can know if the generated action had a good or a bad result. Normally we reward good deeds like defeating the enemy and penalize bad actions like crashing into a wall. However, if we take this traditional approach, the bot will be trained to be extremely efficient, always taking the best actions and becoming a formidable enemy. This work proposes the creation of a bot that behaves like a human and for that we will reward the bot when it takes the actions of the provenance, whether these actions are good or bad, so it will learn to behave exactly as the player behaved.

GAIL does not interact with the expert during training. Instead it GAIL explores randomly to determine which actions bring a policy's occupancy measure closer to the expert's, without requiring the expert to take the action that must be taken by the agent at a given moment of uncertainty in the face of some previously unseen situation. The created brain will act according with the learned policy and behavior coming from the human trainer.

IV. RESULTS

For the validation of our model 7 students were invited to play DodgeBall game in 6 different scenarios. In scenarios A, B and C we generate brains using the traditional reward, that is, at the time of training we give rewards for successful tasks and penalties for unsuccessful tasks. In scenarios D, E and F we used our training method, rewarding all the actions of the sessions of real players who were in the provenance.

After training, we upload the brain to the game agents and run the application. If the NPC's current situation has been

TABLE II
RATINGS GIVEN TO THE NPC ACCORDING TO THEIR BELIEVABILITY.

	Scenarios					
	Traditional Training			Presented Model Training		
	A	B	C	D	E	F
Player 1	1	3	3	2	3	4
Player 2	1	2	3	2	3	3
Player 3	2	3	3	3	3	4
Player 4	1	3	3	2	3	4
Player 5	1	2	2	2	2	3
Player 6	2	2	3	2	3	4
Player 7	1	2	2	2	3	4

seen before, the brain will perform that action and if the situation has not been seen previously the brain will look for the action that most closely resembles the current situation. We ask the players to rate the NPC according to their level of perception as to the similarity of their actions compared to a real player. The rating requested was to give a rating from 1 to 5, with rating 1 being the NPC not acting anything like a player and rating 5 being very similar to a real human player. Table II shows the ratings given by each player in the specified scenarios.

A. Scenario A: NPC trained traditionally with approximately 2400000 steps

The NPC's moves were quite simple, mostly being executed without much intelligence. The initial result of the movement was unsatisfactory, being far from ideal and clearly requiring improvements. The NPC took the balls to make the throw but soon after taking the ball they threw it without aiming. The NPC also did not perform the dash move.

Players found the NPC's moves very simple and defeated him extremely easily. The ball shots were executed without much sense. All seven players did not have the impression that they were facing a real player.

B. Scenario B: NPC trained traditionally with approximately 3400000 steps

Many improvements were observed in the NPCs' movements, with the character moving between obstacles and looking for balls to collect and throw. Compared to scenario A, the result became more satisfactory, especially related to the movements, even using the dash movement. Throwing has improved as well.

Players found the NPC's moves much more similar to what a real player would do, being smart when looking for balls. Players felt that the NPC performed some ball throws for no reason, far from the direction the player was facing. However, around 30% of the ball shots were successful, which had not happened in the previous scenario.

C. Scenario C: NPC trained traditionally with approximately 4500000 steps

Now the NPC's movement is even more efficient, collecting balls quickly and moving between barriers without any collisions. Now the throwing of the balls has improved a lot,

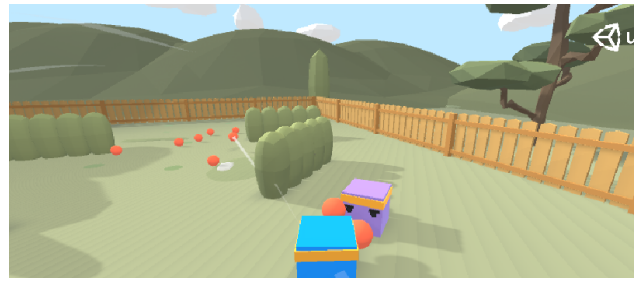


Fig. 6. NPC in scenario C.

We increased the hit rate to about 50%. The NPC in this scenario was considered difficult to defeat and even won 2 matches. In general, players had some impression of being facing a professional player, due to the difficulty. This scenario is illustrated in Fig. 6 that shows that the NPC already has a ball in his hand and is facing the player just before the ball is thrown and hit the player. Despite the improvements some players told us that the NPC was very hard and so it was difficult to believe to be a real player.

D. Scenario D: NPC trained with our model with approximately 2400000 steps

The NPC moved basic without dash and crashed a few times against barriers and wall, but tried to collect the balls. The throwing of the balls was of little precision, with few hits. Players felt a low difficulty when facing the NPC but felt an improvement over scenario A.

E. Scenario E: NPC trained with our model with approximately 3400000 steps

In this scenario the NPC would collect the balls, rarely hit walls and barriers and throw with improved accuracy, but throwing carelessly. Players commented that the NPC was more challenging and appeared to be a player, except when throwing the ball right after collecting it from the ground.

F. Scenario F: NPC trained with our model with approximately 4500000 steps

In the last scenario the NPC moved quite well, using dash to collect the balls and only hit barriers or walls after making a dash. Serious throwing errors have decreased, leaving only the shoots that can be understood as mistakes and would be common for real players to make. Players found this version very fun to face because despite needing improvements it was the version that most resembled a real player. The movement was praised as it moved well but made some small movement errors and some small shooting errors.

G. Analysis of results

According to Table I, In both training models, the greater the number of steps, there were some improvements. The movement became more similar to what a player would do, and ball shots were executed with greater precision and less randomness. The NPC was getting harder to face, becoming

better at moving and throwing the ball. This improvement made the NPC initially look less like a real player and in the end look more like a real player.

However, when comparing the ratings of NPC of scenario B against the ratings NPC of scenario C, both trained in the traditional model of reward, despite the NPC of scenario C being more difficult, has better shooting accuracy and movement, he became so much more difficult that for the players he looked like a well-trained machine because he made almost no mistakes.

Comparing the same amount of steps in the training proposed in our model, the rating given to the NPC of scenario F had an improvement in the perception of being a real player in relation to the rating given to the NPC of scenario E because it became real but continued to make some minor errors that for the players would be the mistakes a real player would make. The NPC of the F scenario in general managed to pass the greatest feeling of being a real player among all the scenarios.

V. CONCLUSION AND FUTURE WORK

This paper presented a novel model, based on provenance techniques, to create NPCs that would take actions similar to those that a player would take, making the perception of immersion of the player who is watching them be increased.

For this, we collect data from different players through provenance and use this data to train NPCs using the GAIL framework. Not using a specialist to train NPCs is essential, so that the NPC does not have optimized performance, but a performance closer to real players. For the validation of our model the DodgeBall game was satisfactory as it is not a very specific game that has complex actions but it is similar to the most games with simple actions as move and shoot. One limitation of our model is that we have to configure the game to train the NPC so we need to have access to the game's source code.

Seven students were called to play the DodgeBall game in six scenarios: on scenarios A, B and C they faced a NPC trained with the traditional reward model giving rewards for good actions and penalties for bad actions. On the scenarios D, E and F they faced a NPC trained with the reward system from our model giving rewards from the all actions that was obtained from the provenance. They rated the NPC according to its behavior during the game. Comparing the ratings that the players gave in the scenarios we observed a good improvement in the ratings from scenarios A, B and C to scenario D, E and F, which means that our methods converge to believable behaviors. All players gave feedback that our trained NPC became more like a real player and the ratings given reflected this. The proposed model give us good impressions and can be scalable to improve the results, generating an NPC that acts very much like a real player.

For future work we should enhance the training process by using more provenance data from more players and see how the agent behaves. It is also important to train for longer times and with a greater number of scenarios for comparison. We want to test our model in a game with more complex

behaviors and check how it performs. We also want to test the believability of an agent created using the model presented in this work with more formal believability testing techniques present in the literature. Another improvement should be create clusters of profiles of different users, so that the recorded sessions follow some kind of user profile (kids, experienced players, casual gamers, etc.). This would make it possible to choose which type of NPC the player would face. We also want to validate our model in other games and allow the use of the feature of influence from provenance to correlate actions and model more complex behaviors.

REFERENCES

- [1] P. De Haan, D. Jayaraman, and S. Levine, "Causal confusion in imitation learning," in *Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS)*, 2019, pp. 11698–11709.
- [2] S. K. S. Ghasemipour, R. Zemel and S. Gu, "A divergence minimization perspective on imitation learning methods," in *Conference on Robot Learning*, 2020, pp. 1259-1277.
- [3] O. Takayuki, J. Pajarinen, G. Neumann, J. A. Bagnell, P. Abbeel and J. Peters, "An algorithmic perspective on imitation learning," *Foundations and Trends® in Robotics*, vol. 7, no. 1-2, 2018, pp. 1-179.
- [4] A. Billard and D. H. Grollman, "Robot learning by demonstration," *Scholarpedia*, vol. 8, no. 12, pp. 3824, 2013, doi: 10.4249/scholarpedia.3824.
- [5] J. A. Bagnell, "An invitation to imitation," Robotics Institute, Carnegie Mellon University, Pittsburg, KA, USA, tech. report CMU-RI-TR-15-08. 2015.
- [6] A. Billard, S. Calinon and R. Dillmann, "Learning from humans," in *Springer handbook of robotics*, B. Siciliano, O. Khatib, Berlin, Germany: Springer, 2016, pp. 1995–2014.
- [7] S. Higgins, "PREMIS data dictionary for preservation metadata," 2009. [Online]. Available: http://www.rsp.ac.uk/documents/PREMIS_V-2-1-2009-03.pdf
- [8] T. Kohwalter, E. Clua and L. Murta, "Provenance in games," in *Proceedings of the 2012 Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*, 2012, pp. 162–171.
- [9] J. Ho and S. Ermon, "Generative adversarial imitation learning," in *Proceedings of the 29th Conference on Neural Information Processing Systems (NeurIPS)*, 2016, pp. 4565–4573.
- [10] A. Cohen *et al.*, "On the use and misuse of absorbing states in multi-agent reinforcement learning," 2021, *arXiv:2111.05992*.
- [11] J. Togelius, R. D. Nardi and S. M. Lucas, "Towards automatic personalised content creation for racing games," in *2007 IEEE Symposium on Computational Intelligence and Games*, 2007, pp. 252-259.
- [12] M. Miranda, A.A. Sanchez-Ruiz and F. Peinado, "A neuroevolution approach to imitating human-like play in ms. pac-man video game," in *Proceedings of the 3rd Conference of the Spanish Association for Videogames Sciences (CoSECivi)*, 2016, pp. 113-124.
- [13] J. Ortega, N. Shaker, J. Togelius and G. N. Yannakakis, "Imitating human playing styles in super mario bros," *Entertainment Computing*, vol. 4, no. 2, pp. 93-104, 2013.
- [14] M. Miranda, A.A. Sanchez-Ruiz and F. Peinado, "A CBR approach for imitating human playing style in ms. pac-man video game", in *International Conference on Case-Based Reasoning (ICCBR)*, 2018, pp. 292-308.
- [15] M. W. Floyd, A. Davoust and B. Esfandiari, "Considerations for real-time spatially-aware case-based reasoning: A case study in robotic soccer imitation," in *European Conference on Case-Based Reasoning (ECCBR)*, 2008, pp. 195-209.
- [16] I. V. Karpov, J. Schrum and R. Miikkulainen, in "Believable bot navigation via playback of human traces" in *Believable bots*, Berlin, Germany: Springer, 2013, pp. 151-170.
- [17] C. Cruz and J. A. R. Uresti, "HRLB²: A Reinforcement Learning Based Framework for Believable Bots," *Applied Sciences*, vol. 8, no. 12, 2018, Art no. 2453, doi: 10.3390/app8122453.
- [18] C. Pelling and H. Gardner, "Two human-like imitation-learning bots with probabilistic behaviors," in *2019 IEEE Conference on Games (CoG)*, 2019, pp. 1-7.