

# Advancing DRL Agents in Commercial Fighting Games: Training, Integration, and Agent-Human Alignment

Chen Zhang<sup>1</sup>, Qiang He<sup>3</sup>, Yuan Zhou<sup>2</sup>, Elvis S. Liu<sup>\*2</sup>, Hong Wang<sup>2</sup>, Jian Zhao<sup>1</sup>, Yang Wang<sup>\*1,4,5</sup>

<sup>\*</sup> Corresponding author, <sup>1</sup> University of Science and Technology of China (USTC) <sup>2</sup> Tencent Games <sup>3</sup> Institute of Automation, Chinese Academy of Sciences, China

<sup>4</sup> Key Laboratory of Precision and Intelligent Chemistry, USTC <sup>5</sup> Suzhou Institute for Advanced Research, USTC

## Motivation & Insight

Deep Reinforcement Learning (DRL) agents have demonstrated impressive success in a wide range of game genres. However, previous research primarily focuses on optimizing DRL competence rather than addressing the challenge of prolonged player interaction.

To bridge this gap, we proposed a practical DRL agents system *Shūkai* for commercial fighting games. *Shūkai* has been successfully deployed to Naruto Mobile, a popular commercial fighting game deployed by Tencent Games.

To solve the generalization and training efficiency problem caused by the large character pool of Naruto Mobile, we introduce **Heterogeneous League Training (HELT)**. HELT achieves balanced competence, generalizability, and training efficiency by utilizing quantization of network states and employing a heterogeneous league exploiter.

*Shūkai* also implements specific rewards to align the agent's behavior with human expectations. We have developed an expert prior-based evaluation system for assessing semantic-level behavior of both DRL agents and humans.

## Naruto Mobile

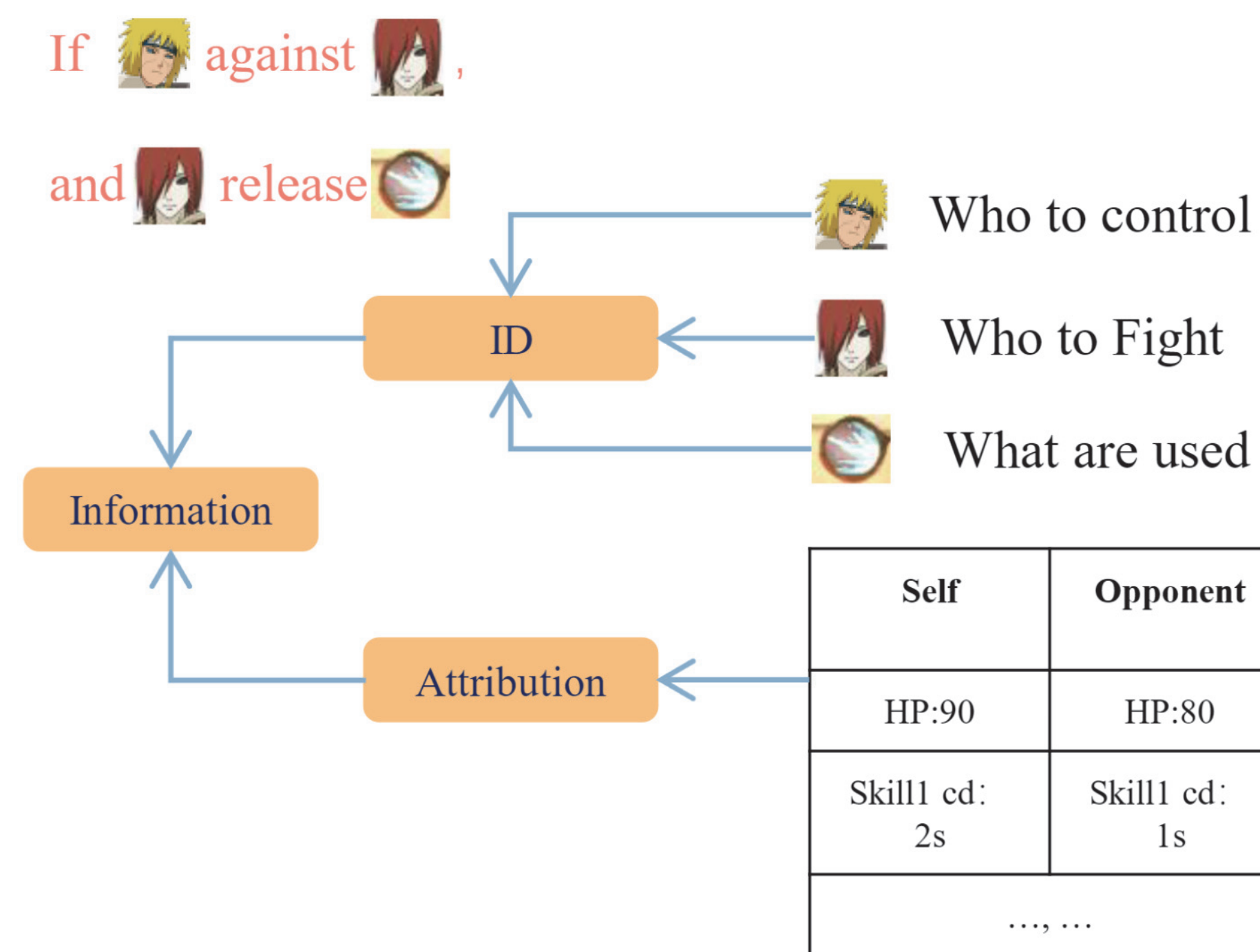


Naruto Mobile is an online fighting game developed by Tencent Games with over 100 million registered users. Naruto Mobile has a large-scale character pool with more than 400 characters (ninjas). Each ninja has its special unique characteristics. Each episode of Naruto Mobile consists of two adversarial ninjas. Players of Naruto Mobile can choose a ninja from the character pool and use the ninja to fight against opponents. The winning condition for all players is to defeat their opponents, and the episode terminates when the condition is satisfied or the timeout.

Naruto Mobile is an online fighting game developed by Tencent Games with over 100 million registered users. Naruto Mobile has a large-scale character pool with more than 400 characters (ninjas). Each ninja has its special unique characteristics. Each episode of Naruto Mobile consists of two adversarial ninjas. Players of Naruto Mobile can choose a ninja from the character pool and use the ninja to fight against opponents. The winning condition for all players is to defeat their opponents, and the episode terminates when the condition is satisfied or the timeout.

## Method

### Heterogeneous Agents



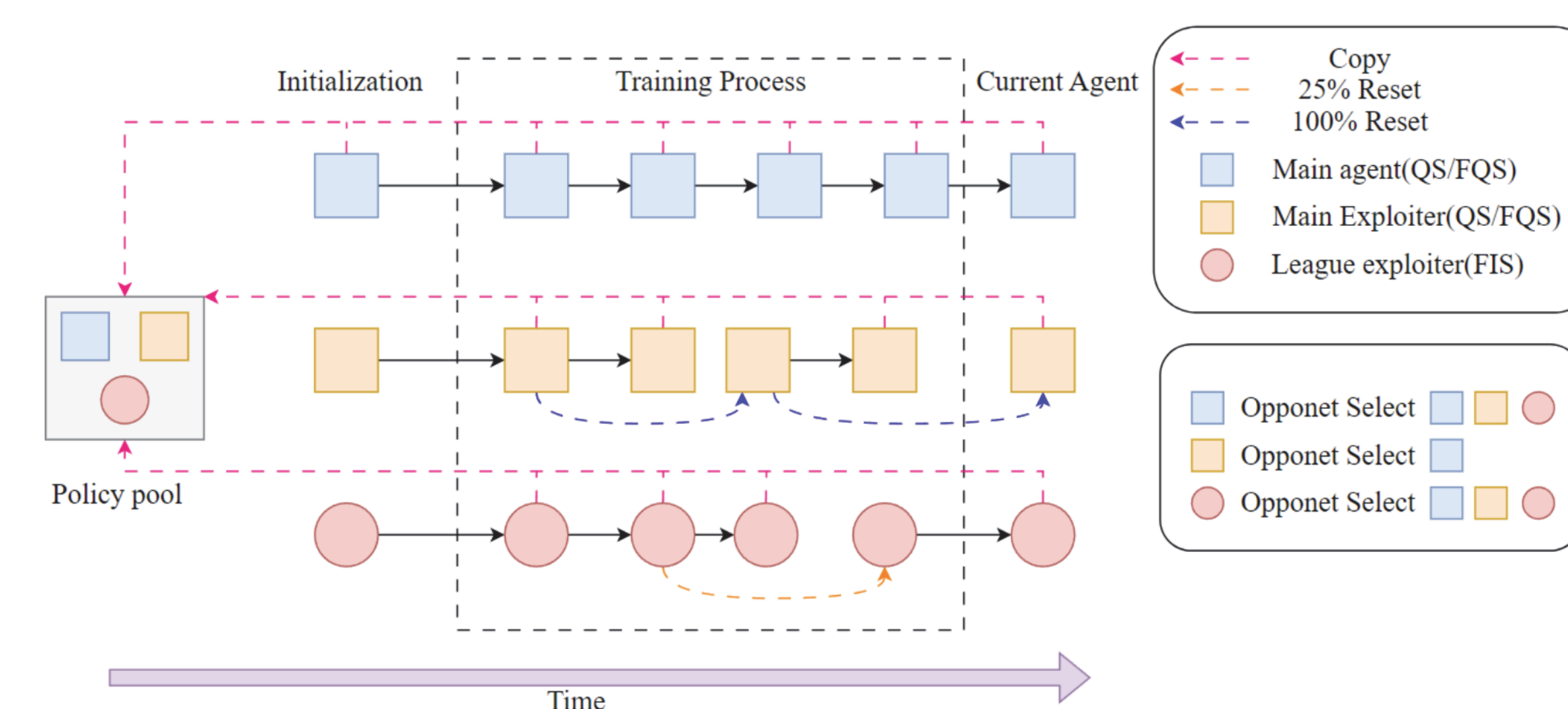
In Naruto Mobile, information can be divided into ID-based info and attribution info. Based on these info, we define three different networks to model self and opponents:

◆ **FIS**: ID + attribution for self and opponents.

◆ **QS**: ID + attribution for self and attribution for opponents.

◆ **FQS**: attribution for self and opponents.

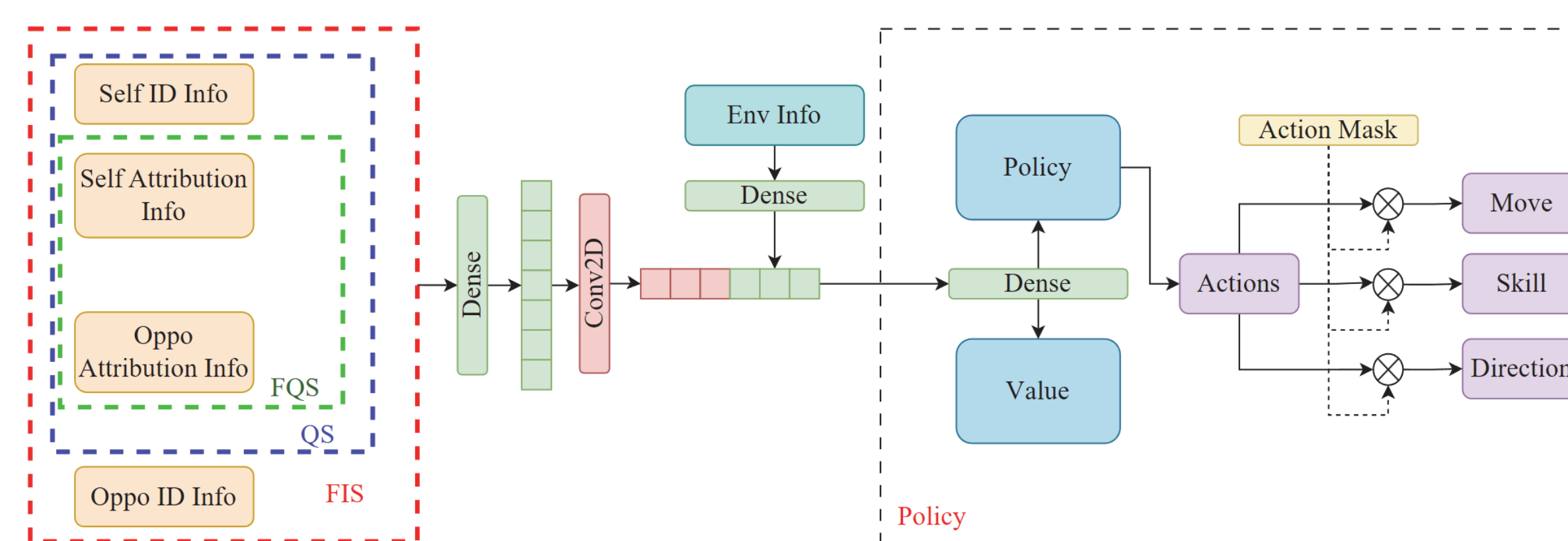
### Heterogeneous League Training



◆ **Main agents**: QS/FQS structure, defeating all opponents in the policy pool, never reset.

◆ **Main exploiters**: same structure with main agent, only fighting contemporary main agents, 100% reset.

◆ **League exploiters**: FIS structure, defeating all opponents in the policy pool, 25% reset.



### Specific Reward for Agent-Human Alignment

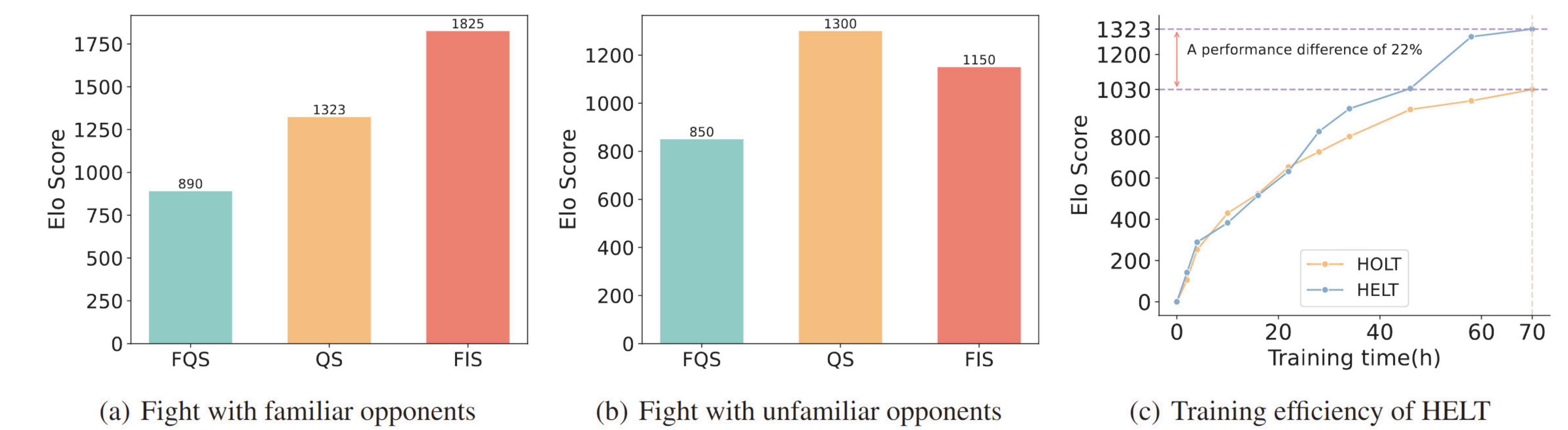
◆ **Balanced**: The agent will trade off the opponent's remaining health point, its health point, and the available resources. It aims to strategically defeat the opponent at the right moment and in a suitable manner.

◆ **Aggressive**: The agent will focus on the opponent's remaining health points. The objective of the agent is to defeat the opponent as quickly as possible.

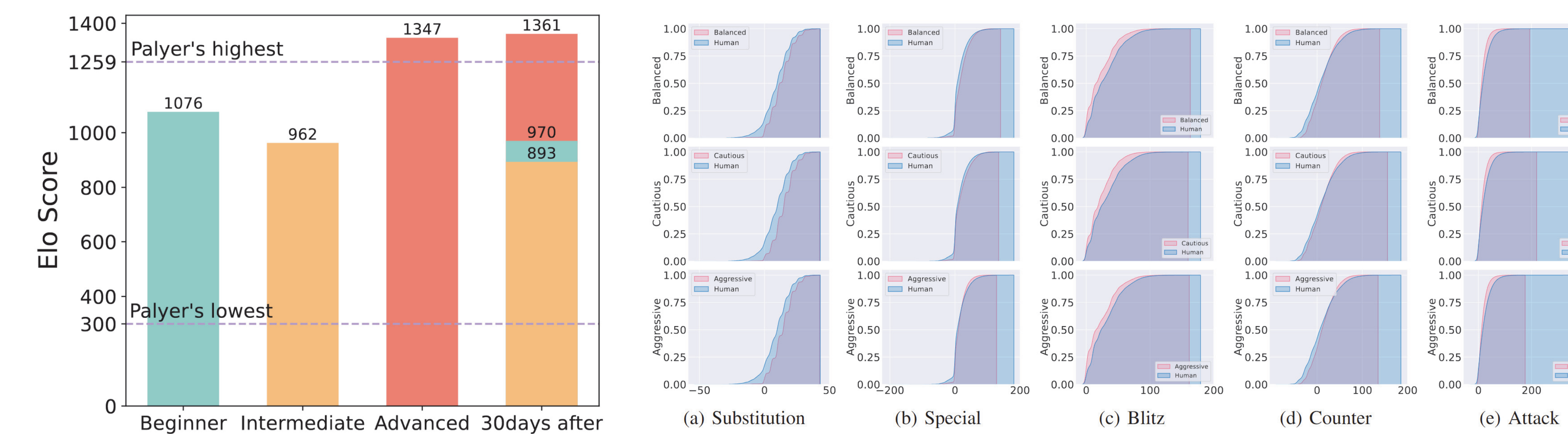
◆ **Cautious**: The agent will prioritize its resources and remaining health points, aiming to minimize damage taken while winning the game.

## Experiment

### Competence, Generalizability, and Training Efficiency



### Human Evaluation and Agent-Human Alignment



◆ During the training process, *Shūkai* was trained using **13%** of the characters (totaling over 400), but demonstrated **adaptability to all characters**.

◆ *Shūkai* has been deployed in Naruto Mobile for over a year, with a total of over **2.1 billion matches played**. It has provided a variety of activities for the game and served as a valuable training partner, helping players improve their skills.

## More Details

Paper



Email



Contact us

