Internship Proposal: Reinforcement Learning from Human Feedback (RLHF) for Human–AI Collaboration

Context

Recent advances in artificial intelligence (AI) have led to systems that surpass human experts in competitive domains such as Chess, Go, Dota 2, and StarCraft II. However, building AI agents that can cooperate effectively with humans remains a major scientific challenge. Most AI training still relies on self-play resulting in policies that perform poorly when paired with real humans. At ISAE-SUPAERO, the HAICO project (Human–AI Collaboration) investigates how agents can adapt to human behavior, interpret implicit cues (e.g., gaze, intention), and communicate their intentions to AI for more natural and fluent teamwork. The environment used is a cooperative cooking game inspired by Overcooked, implemented in Unity game engine.

Objective Project & Environment description

The goal of this internship is to explore Reinforcement Learning from Human Feedback (RLHF) in a cooperative setting. Instead of learning only from environment rewards, the agent's policy will be shaped by a human partner's feedback during collaboration. Specifically, the human teammate will be able to assign positive or negative rewards based on how helpful, efficient, or intuitive the agent's actions feel during the joint task. The intern will investigate how such feedback influences learning stability, team performance, and perceived fluency.

Objectives

In detail, the objectives of the internship are to integrate into the existing Overcooked environment (Unity) an interface that enables real-time human feedback via a keyboard, controller, or graphical interface. The intern will ensure the synchronization of gameplay, human inputs, and reward logging, and will then fine-tune the agent's policy using human feedback as either a direct reward signal or a modeled one, through reward shaping or a learned reward model. The work will also involve investigating the stability of training and the relative weighting between environment-based and human-based rewards. Finally, the intern will collect and analyze data from human–agent collaboration sessions, evaluate performance, fluency, and behavioral alignment metrics, and compare RLHF trained agents with baseline agents fine-tuned without human feedback.

Profile of the Ideal Candidate:

Educational Background

An MSc or Engineering student with a robust background in Computer Science, Robotics, Cognitive Science, AI, and Reinforcement Learning is preferred. Familiarity with human behavior, neuroscience and cognitive sciences will be a notable plus.

Figure 1: Overcooked environment developed with Unity

Technical Skills

Candidates should showcase scientific aptitude, programming proficiency in languages like C#, Python, or C++, and familiarity with the Unity game engine. However, those keen to learn coding will also be considered.

Personal Attributes

A passion for cutting-edge research, the capability to work independently, adaptability, and exemplary communication skills are essential.

Administrative information

Laboratory location

DCAS Department
ISAE-SUPAERO (Institut Supérieur de l'Aéronautique et de l'Espace)
10 Av. Edouard Belin, 31400 Toulouse

Compensation

Interns will receive a stipend of 4.35€ net per hour, equating to approximately 609€ per month.

Duration

Six months, commencing in February or March 2026, with potential opportunities for a subsequent PhD.

Application to be sent to supervisors:

Prospective candidates should forward their CV and cover letter to:
Dr. Christophe Lounis at `christophe.lounis@isae-supaero.fr`
Dr. Caroline Chanel at `caroline.chanel@isae-supaero.fr`.