# Better Understanding of Humans for Cooperative AI through Clustering

**by Edward Su**

Thesis submitted in fulfilment of the requirements for the degree of

**Master of Science (Research) in Computing Sciences**

under the supervision of

Dr William Raffe
Dr Luke Mathieson
Dr YuKai Wang

University of Technology Sydney
Faculty of Enginering and Information Technology
March 2024

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Edward Su, declare that this thesis is submitted in fulfilment of the requirements for the award of Master of Science (Research) in Computing Sciences in the School of Computer Science, Faculty of Engineering and IT at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature Production Note: Signature removed prior to publication. Date: 11/11/24

## Acknowledgement

This thesis was only made possible as a result of the consistant efforts of my supervisors, Dr. William Raffe and Dr. Luke Mathieson of the School of Computer Science, Faculty of Engineering and IT, University of Technology Sydney. This was through constant meetings which not only served as a medium of teaching and guidance, but also one of encouragement and sometimes entertainment.

Furthermore, I would like to thank Dr. Bryan Paton of the School of Psychological Sciences, University of Newcastle, for his speedy assistance with addressing the statistical intricacies of the thesis. This saved many weeks of work which would have significantly lowered the quality of the submission.

I would also like to thank my candidature stage examiners Dr. Nico Pietroni and Dr. Yukai Wang of the School of Computer Science, University of Technology Sydney, who provided insightful feedback which helped guide the direction of the thesis.

Finally, I would like to thak my friends for keeping me entertained throughout the years and to my family for their constant support and encouragement.

# Contents

## List of Figures

# List of Tables

**Abstract**

Cooperative AI and AI alignment research are increasingly important fields of study as machine learning models are becoming more prevalent in society. Applications such as self-driving cars, realistic AI in games and human-AI teams, all require further advancement in cooperative and alignment research before more widespread applications can be achieved. However, research in these fields have typically lagged behind other machine learning applications due to the difficulty of creating models that are robust to and can adapt to novel human partners. We attempt to address this through the creation of a framework that uses Archetypal Analysis, a clustering algorithm that finds extremal 'archetype' points in a dataset and expresses each other point as a convex combination of these archetypes. This framework creates understandable archetypes of players, which a reinforcement learning agent can use to adapt accordingly to unseen partners. We show that this framework not only results in performance comparable to other cooperative benchmark models, but also achieves higher levels of perceived cooperativeness without the need for human involvement during the training process. As such, we demonstrate that using clustering techniques to better model different types of human behaviour and strategies, can be an effective approach in improving the ability of AI models to adapt to and improve cooperation with novel partners.

# Chapter 1 Introduction

Machine learning research has gone through significant advancement in recent years and, in combination with rapid rises in computing power, has led to widespread adoption of related technologies across diverse industries, fundamentally altering the methodologies employed for task execution and problem resolution. Large language models, exemplified by architectures like Generative Pre-trained Transformers (GPT), have garnered significant attention due to their ability to understand and generate human-like text, finding wide usage in applications such as chat-bots and virtual assistants [190, 24]. Deep learning models such as convolutional neural networks (CNNs) have found success in tasks like object detection, image classification, and semantic segmentation, seeing use increasing use in agriculture for crop monitoring [1, 91], in retail for customer behavior analysis [184], and in healthcare for disease diagnosis [39, 87]. Furthermore, reinforcement learning (RL), a branch of machine learning concerned with decision-making and sequential tasks, has gained prominence for its ability to learn optimal strategies through continual interaction with an environment. This paradigm has been used to produce models capable of competing with humans in high skill games such as Chess, Go and Dota 2 [33, 158, 129].

Despite large progress in machine learning research, there exists a conspicuous gap in exploring the development of AI models capable of cooperating with humans. As these technologies become increasingly complex and integrated into society such as the advent of self-driving vehicles [54, 112, 174] and human-AI teams [5, 9, 18, 19], evidence shows that more work needs to be done on improving their capacity to cooperate with humans. This is not a simple task and cannot be easily addressed by minor changes in the framework of AI models, but rather, require specific attention into developing systems where the dynamics of teamwork, trust and commitment are prevalent [50]. This includes tasks such as communication between one another, distributing responsibilities and adapting to preferences, all of which are largely understudied in comparison to other fields within machine learning.

Early attempts at developing AI capable of cooperation took the form of multi-agent reinforcement learning (MARL), where multiple RL agents are trained simultaneously with one another [127, 187, 90]. In this approach, agents attempt to maximise a reward function similar to traditional RL, but have the additional property that agents are able to interact with one another and are able to share knowledge, communicate, and perform joint actions [157]. MARL has found success in numerous areas spanning tasks in games [17], economic interactions [134], and joint-decision making [40], but these have largely been between AI agents. It is often the case that RL models generalise poorly to human partners in cooperative settings [32]. As a result there has been more demand for human involvement in the training process.

Approaches that integrate humans during training generally do so in order to enhance the speed in which AI models converge to optimal policies, avoiding the time-consuming process of exploring actions that, with human intuition, are obviously bad. In Human-in-the-Loop Reinforcement Learning (HRL) and Interactive Policy Learning, human feedback is leveraged to guide the learning process and expedite training, often by having a human oversee the actions of an AI model and rating them as they occur [41, 104, 8, 64, 79]. These approaches are, however, very expensive and require significant human effort and expertise. Though advancements have been made to reduce the amount of data required and enhance the scalability of these algorithms, it has yet to see popular adoption [6]. Furthermore, it can be challenging to gather reliable human feedback as humans may provide inconsistent or conflicting feedback based on subjective preferences, biases, or misunderstandings. In many cases, these issues can exacerbate problems in the training process leading to worse outcomes than training without human input.

An alternative approach is the use of techniques that aim to replicate human behaviour as closely as possible, such as imitation learning, behaviour cloning, and inverse reinforcement learning, which attempt to replicate human action as closely as possible with certain distinctions between them [32, 67, 58, 84]. In imitation learning, an agent learns a policy by observing and mimicking the behavior of a human expert. This approach involves training the agent to map states or observations to actions, often using supervised learning techniques. Similarly, behavior cloning involves directly copying the actions taken by a human demonstrator in a given state, without explicitly modeling the underlying decision-making process. Both imitation learning and behavior cloning are effective for tasks where human expertise or demonstrations are readily available but may struggle in environments with high variability or sparse rewards. In contrast, inverse reinforcement learning (IRL) involves inferring the underlying reward function or preferences of a human demonstrator from observed behavior [7, 79]. By learning the implicit goals or objectives driving human behavior, IRL enables agents to generalize beyond specific demonstrations and adapt to new situations. These methods have found success when trained with and generally improve the ability for AI models to adapt to humans more effectively. However, it is unclear whether this is due to having a stronger understanding of human partners or due to the improved robustness derived from exposure to a greater variety of behaviours [32].

A set of techniques that has seen significant applications for better understanding and modelling data across industries are clustering algorithms [130, 55, 132]. Clustering algorithms are a set of techniques that identify underlying structures or patterns within datasets, often segmentating the data into distinct clusters or groups. In doing so, they provide valuable insights into the inherent relationships and similarities among individual data points. Despite the proliferation of these algorithms in many other applications, they have yet to be applied in the context of developing mental models of humans in the context of cooperative agents, using patterns of decision-making, personality traits, or learning styles among individuals. Due to their flexibility of use and practicality of use, they also demonstrate the potential to be used in conjunction with RL models in order to improve collaboration with humans. However, this has also yet to be done and as such, more research needs to be conducted to explore their benefits, costs and ease of integration.

## 1.1 Motivation

Clustering algorithms have long played a pivotal role in various applications across many industries to identify underlying patterns and extract meaningful insights from large datasets [2, 167, 22, 97]. For example, Netflix, one of the largest video streaming platforms in the world, heavily relies on clustering algorithms to support their recommender systems [56]. These systems analyze data points such as user preferences, viewing history, and interactions to develop models of users to personalize the content displayed to users, enhancing content discovery and user satisfaction [131, 183, 189]. Other use cases include e-commerce platforms analyzing user behaviour to predict purchasing patterns to launch targeted promotions, as well as social media platforms identifying communities of users using information of shared interests to facilitate more engagement [106, 113, 55]. In the realm of video games, clustering algorithms are increasingly utilized for modeling player behavior and preferences to deliver immersive and personalized gaming experiences [167, 170, 62, 61, 60]. By analyzing player actions within a game environment, clustering techniques can identify distinct player profiles that can be leveraged to achieve various outcomes such as dynamically adjusting game mechanics and difficulty levels. Despite the wide applications of clustering algorithms, they have yet to be used in the context of SMMs to improve cooperation in human-AI teams [36, 67] even in spite of their proven capabilities in modelling human personalities and creating user profiles.

Though applications of human profiles in the context of RL have been minimal, there has been much research into the features that make up personalities, a crucial piece of knowledge when determining what techniques are best for modelling profiles. The main approaches to model these personalities are the Five-Factor model [15, 180, 181] and the HEXACO model [11, 13, 12], which though they differ from one another in the choice of traits, each express human personality as a combination of various extremal traits or features, rather than a single, fixed trait. Using the Five-Factor model as an example, it expresses personality as a mixture of the traits: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. Within these traits exists a scale between the extremities of the given trait. For example, extraversion on one end of the scale describes extroverted people who are characterized by seeking out social interactions and external stimulation while the other end of the scale describes introverted people who typically avoid social interactions and seek internal stimulation. The result is that when the model is applied to individuals, they will have a unique combination of different degrees of these traits which when combined, define their characteristics such as how they approach learning and process information [26]. Insight from these models has proven that to capture the complexity of human behavior and preferences accurately, it is necessary to consider the interaction and interplay of multiple traits and factors.

However, how these traits manifest in behaviour can differ significantly between contexts. In practice, this can be difficult to discern. Various clustering techniques are utilized to model human profiles in research. Latent Profile Analysis (LPA) [150] is employed to extract homogeneous clusters based on common response profiles, with studies suggesting an optimal number of clusters for meaningful insights. This allows for more meaningful mixtures of classes. Cluster analysis procedures, such as hierarchical and non-hierarchical-k-means, have been used to identify coping profiles among individuals, revealing distinct coping strategies and their impact on stress and health-related behaviors [59]. An area which has seen significant applications of clustering techniques are games, which involves the use of game metrics such as the number of times a certain action is performed or game completion time. These are simple points of data but can still support developers in adapting to their needs during gameplay such as providing hints or support when the player is struggling or for future products where data points on player data can inform future game designs [60]. In a few cases, they have also been used to help AI agents within the virtual environment to adapt to player behaviour by finding trends in movement and combat [60]. These methods often use unsupervised clustering techniques such as neural gas [149], bayesian networks [169], and emergent self-organizing maps (ESOMs) [99], which have found great success in finding underlying trends in the data. This is due to avoiding the need for manual specification of player types and allowing the technique to determine optimal clusters. These techniques will be explored in more detail in chapter 2 of the thesis.

Games are a popular medium for learning more about player personalities, which is due to how they mirror many aspects of real-world interactions. Players often need to communicate, strategize, and adapt to dynamic scenarios with one another to achieve common goals. In addition, games offer a controlled setting where individual factors can be examined in isolation as rules and goals can be changed with significantly less effort than environments set in reality, which are often more complex and less predictable. The use of games as test beds also allows for efficient collection of large volumes of relevant player data such as actions they perform and the time spent doing certain tasks. By using games as a test bed, researchers can iteratively improve AI systems in an iterative and scalable manner, without worrying about the unpredictability of real-world settings.

These techniques, however, either classify profiles as a single class or mixtures of multiple average classes which do not capture the frameworks describing human profiles as mixtures of extremal traits. A lesser-known algorithm that meets these requirements is Archetypal Analysis (AA) [47], which, in contrast with other model-based clustering algorithms, groups data points as convex combinations of archetypal points rather than mean points. This unique property of AA allows representations of humans to be a mixture of various extremal behaviour traits as opposed to any single one or a mixture of mean traits, which better aligns with our understanding of human personality. As an example, assume that in a dataset of player behaviour, AA found 3 archetypal points, which describe supportive, leadership, and learner traits. In this scenario, every existing entry in the data set, as well as future entries, could be expressed as a combination of these points, such as 0.6 supportive, 0.3 leadership, and 0.1 learners. While archetypal analysis has seen some use in several research fields [51, 152, 22, 69, 97], it has yet to been used in the context of classifying player personalities to base action adaptations on and better cooperate with them. The work in this thesis explores the steps taken to achieve this goal.

## 1.2 Aims

To address the gaps of knowledge mentioned previously, this thesis aims to explore potential applications of AA for constructing mental models of humans that are flexible, and do not require bespoke architecture. In doing so, it hopes to advance the capability for AI models to cooperate with human partners and provide a foundation for future research into cooperative-AI research. This can be broken down into the following aims:

1. To test whether the use of AA for the classification of player strategy and, subsequently, the construction of player models will improve AI alignment and coordination with human players.

2. To investigate methods of measuring AI alignment with human players to allow for better assessment of agent cooperation.

## 1.3 Research Questions

In order to address the proposed aims of the paper, the following research questions have been formulated.

1. (RQ1) Is archetypal analysis suitable for use in a cooperative RL agent?
    (a) (RQ1 - A) Is archetypal analysis effective in classifying a human partner's playstyle?
    (b) (RQ1 - B) What measures are required to successfully integrate archetypal analysis for use in a cooperative RL agent?
    (c) (RQ1 - C) Does formulating a better understanding of human partners using clustering algorithms enable for higher levels of perceived cooperation and or team performance?
    (d) (RQ1 - D) What are the limitations of clustering algorithms for improving AI alignment with humans and what are the implications for AI safety?
2. (RQ2) How can AI alignment with human player personalities be evaluated?
    (a) (RQ2 - A) What existing methods are capable of measuring AI alignment with human players?
    (b) (RQ2 - B) How do these methods assess agent cooperation and alignment with human player goals?
    (c) (RQ2 - C) Are there any qualitative techniques that can help compare levels of alignment between AI agents and human players?
    (d) (RQ2 - D) What are the limitations of these techniques in evaluating AI alignment with human players?

## 1.4 Objectives

In order to address the research questions presented, the following objectives were established.

1. Using a multi-agent environment that facilitates diverse playstyles and strategies, train a variety of AI agents to complete tasks within them.
   (a) Generate playthrough data in the environment and conduct archetypal analysis (AA) on it to represent potential playstyles. (RQ1 - A)
   (b) Develop a custom ensemble RL agent that takes advantage of AA to better adapt to partners. (RQ1 - B)
   (c) Conduct experiments that evaluate the performance of a custom agent that takes advantage of archetypal analysis, compared with benchmark agents developed by Carroll et al. [32]. (RQ1 - C,D)
2. Apply techniques used for measuring AI cooperativity with humans in an experiment where participants partner with AI models in the aforementioned environment.
   (a) Review existing literature on methods used for measuring AI cooperativity with humans.
   (b) Test how effective different approaches are for measuring cooperativity by applying them to the test environment referred to in Objective 1 in an experiment. (RQ2 - B,D)
   (c) Examine how, within these approaches, AI models could be compared to one another to determine which agent is the most cooperative. (RQ2 - C,D)

## 1.5 Out of Scope

The nature of cooperation has been a significant subject of discussion across many fields ranging from psychology, the social sciences and artificial intelligence [141, 171, 50, 14, 142]. Tuomela [171] proposes a definition of cooperation as a product of involved agents having mutual goals, which in turn, induces some form of joint action to achieve through an agreement, whether implicit or explicit. This contrasts with alignment, which is the scenario where one agent does what another agent wants it to do [31]. The distinction often made is that in cooperation, both agents have some goal of their own, while in alignment, one agent does not. For this thesis, we make no guarantees that the agents developed would necessarily fall under the category of cooperative models but define them as cooperative models nonetheless to make their intentions clear to the general audience. More specifically, this thesis focuses on the objective of AI alignment with human players.

In addition, cooperative AI agents are typically considered agents that can cooperate to solve nontrivial problems involving trust, deception, and commitment [50]. However, to adequately implement a test bed and procedure that sufficiently facilitates these features is difficult and time-consuming with much work in this area being still in its infancy [36, 42, 30, 44, 103]. Training models capable of achieving those outcomes also presents a significant challenge due to the widely expanded state space and memory requirements. It is for these reasons that implementing a complete cooperative agent capable of these features is deemed out of scope.

In many cooperative environments, hidden information is often a factor such as in the cooperative board game Hanabi [20]. In these settings, players are faced with situations where they do not have perfect information to make informed decisions about what action to pursue. This often requires players to communicate with one another to achieve reasonable results. In this thesis, we deem explicit communication between humans and AI out of scope for the thesis as we already handle communication of intent implicitly using our custom approach.

## 1.6 Significance

This thesis contributes to the field of cooperative AI in three main areas.

1. Firstly, the development and testing of AA for use in RL to create AI agents capable of cooperating with human partners. We will refer to this as the "AA agent" for the remainder of the paper. As mentioned previously, the use of clustering algorithms for improving the ability of RL models to cooperate with human partners is largely unexplored and there is little research done on the development of AI agents which are capable of adapting to unique human behaviour. Thus, the development and testing of these algorithms would be a valuable contribution to the broader, growing field of cooperative AI [49], and more specifically in the context of games.

2. Secondly, an experiment that demonstrates the relevance of perceived cooperation when examining the success of cooperative AI agents, and their relative performance. Most research conducted in cooperative AI evaluates the effectiveness of AI agents in cooperating with human players through a quantitative score representing the overall performance of the human-AI team in completing some task [20] [32]. This ignores the mental aspects of cooperation which have significant impacts on the quality of cooperation, such as how humans perceive or trust the AI agent to supplement or direct a strategy [9]. This thesis aims to formulate an experiment that considers and compares the quality of cooperation exhibited by AI agents, in addition to objective performance measures.

3. Finally, informed by our development of the AA agent and the results of our experiment, we provide suggestions for future directions to take

research in the field of cooperative AI. This includes improvements that could be made to the AA agent as well as to the approach we took during our experiment to compare levels of cooperation between different AI agents and human partners.

## 1.7 Thesis Structure Overview

There are eight chapters in this thesis, which address the research aims and objectives described previously, and the research performed to achieve them. Chapter 2 provides background knowledge of relevant work that has been done to train cooperative agents using reinforcement learning, and modelling human personalities. Chapter 3 continues to explore previous work but with the goal of providing a broad view of techniques in the field rather than a detailed explanation. After this, there are two chapters that cover the experiments: Chapter 4 describes how the Archetypal Analysis was used in conjunction with RL to create a AA agent capable of cooperating with human partners, and a description of the environment it was trained in. Furthermore, it discusses the use of additional qualitative data as a metric for evaluating the cooperativity of RL agents. Afterwards, Chapter 5 outlines the structure of the experiment that was conducted to evaluate the products of Chapter 4. The results of this experiment is provided in Chapter 6 and their implications and significance are discussed in Chapter 7. The thesis concludes in Chapter 8 which reviews the work that was done and summarises the answers to the research questions. A discussion of the limitations of the work and potential directions for future work are then supplied. Afterwards, there is a small appendix of additional supporting figures and a bibliography with all the references used throughout the thesis.

# Chapter 2 Background

This heading covers concepts that make up the core of, and which are essential for the understanding of our research into cooperative AI models. It begins with a description of reinforcement learning as well as related topics in the form of multi-agent reinforcement learning and self-play. Next, an explanation of the Archetypal Analysis clustering algorithm as well as descriptions of recent configurations is presented. Finally a discussion of how ensemble learning functions as well as insights into potential applications will be provided.

## 2.1 Training Cooperative Agents

Research into AI agents capable of cooperating with humans involves exploring how agents can determine common goals and execute strategies during uncertainty. This is however difficult to achieve without some threshold of competency in completing a task without the involvement of a human partner.

A common technique used to achieve this is Reinforcement Learning (RL) which has seen great success in 1v1 zero-sum games such as Go and Chess [33, 158, 129], as well as some partially cooperative settings such as DOTA 2 [129]. RL is a prominent field of machine learning that centers around developing AI agents that learn optimal decisions through interactions with an environment [163]. This is achieved by providing feedback to RL agents, rewarding them when they perform actions that progress them toward the desired goal, and punishing them when they do not. These agents attempt to maximize the reward they receive from the environment then tune their actions appropriately and improve in performance over time. This process is visualised in Figure 1. A common approach to modelling this process is through a Markov Decision Process (MDP) where after performing some action and receiving information about the proceeding reward and state, they select a new action in response.

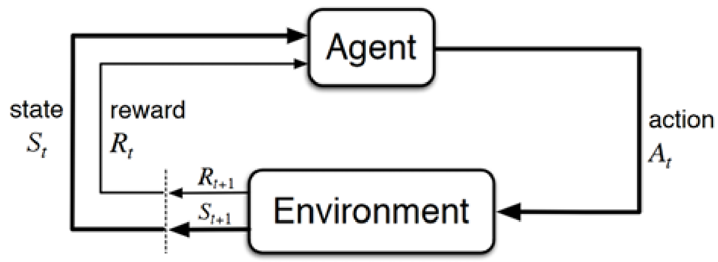


Figure 1: A visual representation of reinforcement learning [163]

A core challenge in the field of RL is the exploration, and exploitation trade-off which describes how the agent is unaware of the true reward of any given action until it performs the action at some point. This is made further difficult as a result of environments where actions provide stochastic rewards and or rewards that have temporal properties.

A more specific problem RL faces in training agents capable of cooperation is in modelling multiple agents within an environment. Consider a scenario where two agents are learning simultaneously, treating the other as part of the environment. Both agents would react to the environment as well as the actions of the other agent. This causes a cycle of continuous adaptation to the policy of the other agent and presents a dynamic learning problem that makes it challenging to learn and converge at any given policy. This problem of convergence is a key focus in the subset of RL known as Multi-agent Reinforcement Learning (MARL) and describes how agents learn and ultimately reach an optimal, stationary policy where there is no longer any benefit in further adjustment [187].

A potential solution to this problem is a centralized approach that has all agents within the environment learn a single policy that dictates the actions of every agent [187]. This approach has agents attempt to maximize a reward function similar to traditional RL, but the state of an environment as it is interpreted by

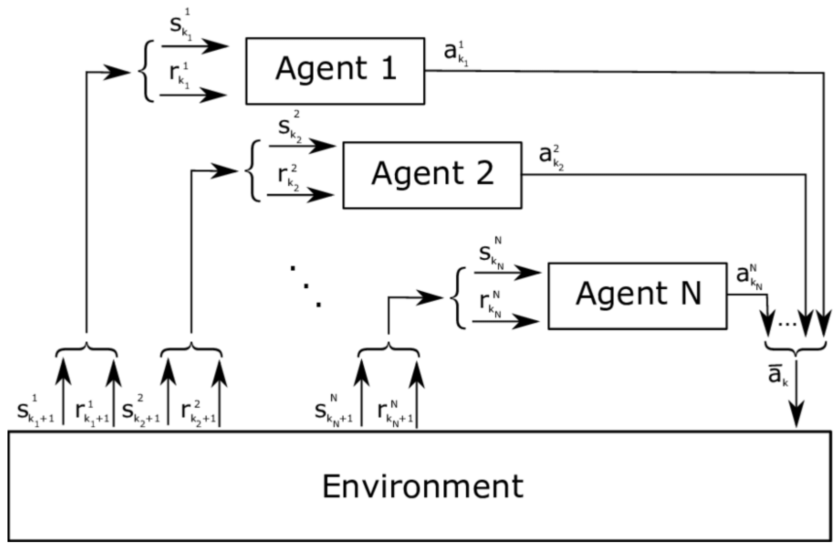


Figure 2: A visual representation of MARL (Self-Learning Power Control in Wireless Sensor Networks)[163]

each agent is considered to output joint actions performed by them. As a result, agents can share knowledge between themselves and the policy guiding them can converge. However, this approach is limited in its ability to maximize the rewards of individual agents independently from one another. A decentralized approach addresses this by considering each agent as a part of the environment and having them all learn at the same time [127] which is visualised in Figure 2 but this approach suffers from the aforementioned issue of convergence which is no longer guaranteed.

A solution to training agents capable of cooperation using MARL is by having agents train with themselves in the form of self-play, where interaction with themselves within an environment can improve their coordination and decision-making policy. In this case, one agent's policy generally remains static and its actions are considered as part of the environment, while the other agent's policy will be dynamic. The dynamic agent learns to more optimally interact with the static agent and updates its policy accordingly. The updated policy is then copied to the static agent and the process is repeated until termination. This approach avoids the aforementioned problem of convergence posed by MARL as well as the need for novel partners for training, relying solely on an agent's own experiences to drive learning. This approach of enabling agents to experience continual self-development has seen much success in many applications such as games, where agents have beaten professional players in environments such as Starcraft, and Go [33, 158].

However, the self-play approach of training agents to coordinate with one another does not translate well to cooperation with human partners [32]. This is due to the self-play agents overfitting their strategies and decisions to one another meaning that though their strategy may perform very well if precisely followed by all the agents involved, any deviation could cause significant drops in performance. One approach to addressing this is through the involvement of humans in the training process.

Another approach is to train AI agents that can act in a way that resembles a human in the form of human proxies. Past experiments have used behaviour cloning techniques to achieve this which have models learn their policy through exposure to demonstrations by a human [32]. A cooperative RL model will then repeatedly train with this human proxy instead of themselves in self-play to improve their policy. Though this has shown promise in improving the ability of AI models to cooperate with unseen human partners, this is not proven to be because they have learned to better adapt to human partners or whether simply the result of having a more robust policy. This uncertainty is best shown in another experiment that avoids the involvement of humans in the training process and instead trains a cooperative RL model with a population of self-play models in a process known as Fictitious Co-Play (FCP) [162]. FCP demonstrated that even without human involvement, by training with a more diverse set of agents, it is possible to improve cooperation with human partners as a result of more robust policies.

Other methods involving humans in the training process take the approach of eliciting their feedback to guide how RL agents learn and the policies they converge to. In Human-in-the-Loop Reinforcement Learning (HRL) and Interactive Policy Learning, human feedback is leveraged to guide the learning process and expedite training, often by having a human oversee the actions of an AI model and rating them as they occur [41, 104, 8, 64, 79]. For example, if an agent performed an action, a human would provide feedback on whether that action was good or bad. This feedback would translate to the rewards the RL agent receives for that action with a good rating providing positive rewards while bad ratings provide negative rewards. These approaches suffer from having to require significant human effort as even simple RL models can require hundreds of thousands of timesteps to train. Furthermore, these approaches also suffer from biases of the humans providing the feedback. In addition, they can suffer from inconsistencies in the feedback they receive. For example, an action deemed good earlier in training by human overseers could then be regarded as bad later even when the state of the environment is the same.

Common future directions listed for improving cooperative RL models involve improving training partners by improving how human-like they are, and/or by diversifying their behaviour, as well as adapting to human partners in test-time [32, 162]. We believe that to do so, a deeper understanding of how human personalities differ from one another is required.

## 2.2 Modelling Human Personality

Research into human personality has been a common topic across various research areas and fields such as sociology and computer science due to the inherent value of better understanding how humans think and feel. This pursuit has a wide range of applications such as improving the health of individuals, better understanding societal structures, and enhancing entertainment [59, 177, 60].

One prominent framework for understanding personality is the Five-Factor Model (FFM) otherwise known as the Five-Factor personality model, which posits that personality traits can be organized into five broad dimensions: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism (often abbreviated as OCEAN or CANOE) as seen in Figure 3 [15, 180, 181]. According to this model, individuals vary in their levels of each trait, leading to a wide range of personality profiles, a feature that has been used to adjust AI models to perform believable human actions [144]. Research within the FFM framework has shown that while individuals may exhibit dominant traits along certain dimensions, they also possess a unique combination of traits that distinguishes them from others. For example, an individual may be high in extraversion but low in conscientiousness, or high in openness to experience but low in neuroticism. This suggests that human personalities are multifaceted and can encompass a mix of different trait combinations. This model has also been adapted to study player personalities and preferences [26] to explore how they relate to various aspects of gameplay, such as game preferences, play styles, and social interactions.

FFM has been widely applied in various fields due to its comprehensive coverage of fundamental personality dimensions. In organizational psychology, the FFM has been extensively used to predict workplace outcomes such as job performance, leadership effectiveness, and job satisfaction. For example, research by Judge et al. (2002) [92] demonstrated that Conscientiousness, one of the Big Five factors, is a strong predictor of job performance across different occupations and job types. Moreover, studies in clinical psychology have applied the FFM to understand personality disorders and psychopathology. For instance, research by Costa and Widiger (2002) [45] explored the relationship between the FFM traits and personality disorders, highlighting the importance of neuroticism in predicting emotional dysregulation and mood disorders. Additionally, the FFM has been employed in cross-cultural research to examine cultural variations in personality structure and values. By providing a robust framework for under-

[Production note: This figure is not included in this digital copy due to copyright restrictions.]

Figure 3: Five Factor Model

standing personality traits, the FFM continues to serve as a valuable tool for researchers and practitioners in diverse fields, offering insights into individual differences and behavior across different populations and contexts [92, 45].

An extension of the FFM framework is the HEXACO model see in Figure 4 which includes the addition of 'honesty-humility' as a sixth factor [11, 13, 12]. This factor captures individual differences in sincerity, fairness, and modesty, which are not fully represented in the FFM dimensions, and thus offers a more comprehensive and culturally universal perspective on personality, particularly concerning moral and ethical behavior.

The HEXACO model of personality has found applications across various domains, including psychology, organizational behavior, and social sciences. In organizational psychology, the HEXACO model has been utilized to predict

[Production note: This figure is not included in this digital copy due to copyright restrictions.]

Figure 4: Hexaco Model

workplace outcomes such as job performance, leadership effectiveness, and organizational citizenship behaviors. For instance, research by Ashton and Lee (2007) [11] demonstrated that Honesty-Humility, a unique factor in the HEXACO model, predicts counterproductive work behaviors, such as theft and absenteeism, above and beyond the traditional Big Five factors. Moreover, studies in social psychology have used the HEXACO model to investigate individual differences in moral judgment and ethical decision-making. For example, research has shown that individuals higher in Honesty-Humility are less likely to engage in unethical behavior [11], such as cheating and deception, compared to those lower in this trait. Additionally, the HEXACO model has been applied in cross-cultural research to explore cultural differences in personality structure and values [12].

Though there are contentions on the exact choices and numbers of the traits in these frameworks representing human personality, it is widely accepted that personality consists of a combination of various extremal traits or features. Though these frameworks have found much use in qualitative work such as in sociology and psychology, it is difficult to use them as they are for training RL models. Namely, data from the combination of the environment and the agent's actions need to be collected and processed to represent different traits.

## 2.3 Clustering Techniques

A set of techniques that have seen significant applications for better understanding and modelling data across industries are clustering algorithms [130, 55, 132]. Clustering algorithms are a set of techniques that identify underlying structures or patterns within datasets, often segmenting the data into distinct clusters or groups. In doing so, they provide valuable insights into the inherent relationships and similarities among individual data points. Despite the proliferation of these algorithms in many other applications, they have yet to be applied in the context of developing mental models of humans using patterns of decision-making, personality traits, or learning styles among individuals. Due to their flexibility of use and practicality of use, they also demonstrate the potential to be used in conjunction with RL models to improve collaboration with humans. However, this has also yet to be done and as such, more research needs to be conducted to explore their benefits, costs, and ease of integration.

Among clustering techniques, there exists a distinction between supervised and unsupervised techniques. In supervised clustering, the algorithm is provided with labeled data, where each data point is associated with a class or category. The algorithm learns to identify patterns in the data based on the labeled examples provided during training and aims to predict the class labels of new, unlabeled data points based on these learned patterns. Far more common are unsupervised clustering techniques which operate on unlabeled data, meaning there are no predefined categories or classes associated with the data points [149, 169, 99]. The algorithm identifies inherent patterns or structures in the data without any prior knowledge or guidance, grouping similar data points based on their intrinsic properties or similarities. The following section will cover a variety of unsupervised approaches that have been used for extracting meaningful personality profiles in the past.

One of the most popular clustering techniques is the k-means algorithm which is a partitioning-base algorithm that aims to split the data into a predetermined number of clusters. It does so by determining mean points in the dataset to be centroids and iteratively assigning data points to clusters based on the proximity to the cluster centroids, which are updated until convergence [110]. A limitation of this technique is that it is unable to express data points as a combination of clusters, with each data point belonging to a single distinct cluster. Despite this, the k-means algorithm has found extensive use for creating profiles of humans in a variety of industries to gain more insights about them and to make resource allocation more efficient. For example, banks in Indonesia have used the clustering technique to profile their customers using data they have collected such as their residential information, gender, and age [165]. In doing so, banks segment their customers into different demographies and behaviour types, improving their ability to customize and target the services they provide to their consumer base such as by prioritizing certain high-value customers and determining individuals that are at higher risk to lend to. Other applications of k-means include in businesses to gain advantages over competitors by better understanding the personalities of customers such as labelling them as careless, careful, or sensible [96] as well as in healthcare where it has been used to identify coping profiles among individuals, revealing distinct coping strategies and their impact on stress and health-related behaviors [59]. This technique however can struggle to handle noisy datasets that include outlier data points as well as those with varying cluster sizes and densities due to the innate assumption that clusters are spherical and have similar densities [107]. Furthermore, k-means is generally less effective when there is little prior knowledge of the dataset as it requires the specification of the number of clusters beforehand [107]. This means that it is typically a better method for affirming assumptions about a dataset rather than for data exploration.

Another clustering technique that has seen use for the construction of human profiles is Hierarchical clustering. Hierarchical clustering algorithms organize data into tree-like structures, known as dendrograms, based on the pairwise distances between data points [73, 43]. This is achieved by merging or splitting the two closest data points or clusters depending on the approach taken. In the Agglomerative hierarchical clustering approach, each data point starts as a singleton cluster, and the closest pairs of clusters are iteratively merged until a single cluster containing all data points is formed [27, 52]. Conversely, Divisive hierarchical clustering begins with all data points in a single cluster and recursively splits them until each data point forms its cluster [76, 182, 145]. The benefit of hierarchical clustering techniques is that the clusters they generate can be much more interpretable compared to other clustering techniques as when visualized using a dendrogram, nested clusters can be identified and provide a better understanding of the relationships between data points. Furthermore, the technique does not assume the number of clusters and shape of the dataset which can make it more suitable for many datasets and for exploring datasets without prior knowledge about its features. Because of these benefits, hierarchical clustering has seen applications in digital systems such as personalization algorithms [155]. In this framework, hierarchical agglomerative clustering is used to cluster tags on resources such as 'design', 'programming', and 'baseball', and create groups of related content, while user profiles were developed based on their interactions with different resources. A recommender system would then be used to bridge user profiles to resources it deems relevant to them through the groups of related content [155]. Hierarchical clustering methods have also seen usage for modelling social media users, grouping users exhibiting similar behaviour using data such as how often they post [160].

Closely related to hierarchical clustering techniques are graph-based algorithms such as Spectral clustering which also partition data into clusters with a graph-like structure [178, 107]. Spectral clustering works by first constructing a similarity graph from the data representing distances between data points. The Laplacian matrix of the graph is then computed, and its eigenvectors corresponding to the smallest eigenvalues, excluding the first constant eigenvector, are obtained. These eigenvectors capture the low-frequency components or global structure of the data, enabling dimensionality reduction. Once this lower dimensional space is constructed, any standard clustering algorithm such as k-means can then be performed. Because of this dimensionality reduction, spectral clustering is efficient for large datasets despite the costly eigenvector computation [28, 25]. This feature has seen spectral clustering applied to large datasets such as datasets by Facebook, to provide recommendations for friends based on an individual's existing social network [166]. Furthermore, it has found applications in the energy industry for customer segmentation, using patterns in consumer power consumption to help optimize resource allocation and energy management strategies [3].

Other approaches include Density-based clustering algorithms such as DBSCAN (Density-Based Spatial Clustering of Applications with Noise), which identify clusters based on regions of high data density [37, 57, 48, 147]. DBSCAN forms clusters by grouping closely packed data points, while distinguishing noise points in sparser regions. Unlike partitioning-based algorithms, density-based methods are robust to outliers and can discover clusters of arbitrary shapes and sizes, as well as density levels in the case of the hierarchical versions known as HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise)[186, 57]. As a result, DBSCAN has found significant usage for extracting meaning from challenging datasets such as radar pulses and road networks [48, 37]. Similar in function is Density Peak Clustering (DPC) which identifies clusters through the process of finding points with high local density and low proximity to points with higher density, known as density peaks by leveraging density and distance information [77, 153, 185]. DPC is less capable of handling larger datasets compared to DBSCAN but has the benefit of being simpler and generally faster [139, 4]. Density-based techniques though popular for use in many large datasets, have only seen light usage for creating user profiles. An example application was the use of DBSCAN for customer segmentation where past data such as the frequency of purchases, recency of purchases and the amount of money consumers spend in transactions are considered to create customer classes [122, 151]. Another instance was the use of DPC for analyzing the viewing patterns of online content by internet users to fuel a recommendation system [179].

Though the clustering techniques mentioned above have found many applications across industries, they largely fall short in modelling personalities as a mixture of traits as mentioned previously. Model-based clustering algorithms assume that the data is generated from a mixture of probability distributions and aim to identify the parameters of these distributions to infer the underlying cluster structure [117, 116]. The Expectation-Maximization (EM) algorithm is a popular model-based clustering method that iteratively estimates the parameters of the mixture model, such as cluster means and covariances, using the observed data [119, 138]. Model-based clustering techniques provide a probabilistic framework for clustering and can handle complex data distributions, making them suitable for applications where the data does not conform to simple geometric shapes. Latent Profile Analysis (LPA) is an example of this which has previously been employed to extract homogeneous clusters based on common response profiles, with studies suggesting an optimal number of clusters for meaningful insights and class mixtures [150]. Furthermore, Gaussian Mixture Models (GMMs) have been previously used to analyze data of player performance in sports to construct player profiles, which describe what roles they are likely to fill in a team as well as what skills differentiated them the most from other players [161].

Model-based clustering algorithms make progress towards the ideal representation of human personality as a mixture of traits but a large number of them still fall short in a single detail - that is, though they do represent people as mixtures of clusters, the clusters are still largely constructed as mean values rather than extremal values. A promising algorithm that fulfills this detail is Archetypal analysis.

## 2.4 Archetypal Analysis

Archetypal Analysis (AA) is a model-based clustering technique that classifies data points as a convex combination of extremal 'archetypal vectors' as opposed to around mean values [47]. These archetypal vectors are defined as extremal points in the data that create a boundary that encapsulates all other observations in the form of a convex hull. In doing so, the basis vectors are significantly different from one another, which provides more meaningful information when contrasting strategies and makes it simpler to interpret the results achieved [168]. The original AA algorithm seen in Equations 1, 2 and 3, works by initializing a set of archetype vectors 'A' randomly, and iteratively optimizing the weights that represent them 'W' using a least squares minimization algorithm, to reduce the squared error of reconstructed points until convergence [47]. This proves problematic when applied to larger data sets as the computational complexity is scaled quadratically by the number of data points.

$$\min_{A,W} \sum_{i=1}^{n} \sum_{j=1}^{k} w_{ij} ||x_i - a_j||^2 \tag{1}$$

subject to:

$$\sum_{j=1}^{k} w_{ij} = 1, \quad \text{for } i = 1, 2, ..., n \tag{2}$$

and

$$w_{ij} \geq 0, \quad \text{for } i = 1, 2, ..., n \text{ and } j = 1, 2, ..., k \tag{3}$$

A configuration by Bauckhage identifies that data contained within the convex hull created by the data set do not contribute to the residual being minimized and so exclude them from the process to create a 'working set'[23]. In doing so, the AA algorithm is capable of clustering larger data sets as the optimization problems are reduced and continue to be reduced over time as more points are removed from the 'working set'. AA has also been expanded recently to have the ability to function with multivariate data through the use of deep learning in the process [97], support more observation types such as probability vectors for real-world applications [152], and to be more fast and efficient by using a novel implementation of the Huber loss function [38]. Though improvements have been made, AA still suffers from being computationally expensive compared to other clustering methods such as DBSCAN and Spectral clustering, a weakness that is exacerbated by larger datasets and higher dimensional spaces.

Despite this, AA has seen use in many real-world applications [105, 35, 111] such as in the creation of performance profiles for sport athletes [66]. In this experiment, player statistics from basketball and soccer leagues were collected and AA was applied to produce several player archetypes for each sport. The result was the extraction of meaningful traits and profiles for different players in the dataset. For example, in the basketball dataset, four archetypes were found: the 'benchwarmer', 'rebounder', 'three-point shooter', and 'offensive' - with players in the dataset being represented by a combination of these archetypes. Similarly, AA has also been applied in the context of games in the form of a large dataset on over 70,000 World of Warcraft players [168]. This experiment demonstrated the ability to create more meaningful and interpretable profiles from the dataset as compared with other standard clustering approaches such as K-means.

# Chapter 3 Literature Review

This heading covers a variety of techniques that are relevant to the study of cooperative AI agents. It begins by exploring various reinforcement learning approaches and the benefits they offer to cooperative AI models. Secondly, a review of algorithms used for training human-like agents is conducted. The section is then concluded with an analysis of techniques that allow for a better understanding of humans for AI-human cooperation.

## 3.1 Training Agents

As with other areas of machine learning, training models to make appropriate decisions is a core part of developing cooperative agents. Methods that do so need to perform well in the environment they were trained but should also remain robust to new scenarios and partners. The current section of the paper will explore common techniques used for this training process as well as branches of research that have developed out of them to enhance their capabilities.

### 3.1.1 Deep Reinforcement Learning (DRL)

Traditional RL models use Q-learning to assess the value of certain actions by predicting the reward of the next state the action transitions them to, otherwise known as a Q-function. This is achieved through a process of learning what actions typically lead to certain states in the form of action-value pairs. This approach however struggles when scaling to complex environments with much larger state spaces as many more action-value pairs need to be learned. An approach to addressing this problem is known as Deep Q-Networks (DQN) or Deep Reinforcement Learning (DRL), which leverages deep learning to approximate the Q-function [163]. This approximation is achieved through a series of deep convolutional neural networks (CNN) to learn high-level representations of the state space and a process of backpropagation to gradually reduce the error of the approximations. As a result, DRL models have found much success in traditionally challenging environments with large state spaces and high dimensional data [121, 72], including those requiring hidden information and large numbers of agents that need to cooperate with one another [78].

However, the integration of deep learning has also made results for models highly varied and difficult to reproduce[81]. Furthermore, like traditional RL, DRL is not sample efficient and requires a significant amount of data to train from - a direct result of needing a large number of tests and mistakes to learn from. Attempts have been made to rectify this, including the use of model-based RL methods which though tend to be less performant due to biases of the environment, are more sample efficient[125].

### 3.1.2 Hierarchical Reinforcement Learning (HRL)

Hierarchical Reinforcement Learning (HRL) is another extension of RL that aims to handle complex environments with larger state spaces. This is achieved by decomposing complex tasks into a hierarchy of decisions at different levels of abstraction so that monolithic calculations do not have to be made at every given time step [53]. This is often achieved through a process of temporal exploitation where tasks are temporally extended and follow their policies until termination. This behaviour naturally creates a hierarchical structure where higher-level managers provide tasks to lower-level managers in a recursive fashion. Managers in this system are only aware of states as they apply to their level and as such will not have access to rewards and information about the goals of the task set for them, as well as how tasks have been addressed in lower levels [175]. By doing so, the complexity of the state space is progressively reduced as tasks trickle downwards in the hierarchy and frequent, low-level decisions can be made efficiently [71].

This presents great potential for use in cooperative AI settings where this hierarchical structure could greatly improve the efficiency of centralized MARL frameworks as well as better model the hierarchical decisions cooperative agents make regarding the needs of the team and themselves. An example of this was its applications to a game of 'Capture the Flag' where each agent had an inner optimizer that maximized expected rewards and an outer optimizer that finds solutions to winning the game as a whole [88]. This multi-timescale structure supported memory and long-term reasoning that made the outer optimizer suitable for developing high-level strategies while the inner optimizer was still able to resolve shorter-term and lower-level tasks.

### 3.1.3 Evolutionary Algorithms

An alternative approach to training agents are evolutionary algorithms (EA), which aim to achieve optimal rewards by iteratively evolving a population over multiple generations with the actions of high-performing agents being imitated [133]. This works by comparing the performance of all agents in a population and exploring and exploiting them at different intervals. Exploration takes place in the form of randomly perturbing values by a certain factor to evaluate whether the new value is superior in achieving a higher score than previous values. In this way, agents that perform tasks together better will survive while weaker versions will eventually die out in favor of the stronger models [89, 83, 148]. Some methods for exploitation include truncation selection, where if an agent is in the bottom threshold of the population, it will copy the weights of an agent in the top threshold of the population, as well as T-Test selection, where an agent uniformly samples another agent, and copies the weights of the sampled agent if it has a higher performing score than itself. As a result of these larger population sizes, EA typically produces a larger variety of agent behaviours compared to RL methods but struggles to learn from actions and process why certain actions produce less benefit than others [98].

Attempts have been made to incorporate EA into DRL algorithms to take advantage of the diversity of behaviours EA produces and to access the powerful gradient-based methods from the latter to strengthen learning [98]. This hybrid approach works by training an RL agent through the experiences developed by EA and then inserting the RL agent back into the EA population occasionally so it can be exposed to gradient information. Coined as Evolutionary Reinforcement Learning (ERL), this approach strengthens RL with the robustness and immunity to sparse rewards that EA methods have, whilst still retaining the low sample complexity that leveraging gradient methods have.

## 3.2 Human Models

Though the methods listed in the previous section have found success in developing AI that performs well in AI-AI environments, these results do not cross over to AI-human scenarios [32, 20]. This is due to agents often expecting their partners to be optimal in their actions, have a reward function that aligns with their own, or struggle to communicate effectively. These problems highlight the need to include a human in the training process so that agents understand how to coordinate with imperfect actions and without opaque strategies. This section explores the common methods used to do so or create AI agents capable of imitating human behaviour.

### 3.2.1 Imitation Learning

Imitation learning works by replicating an observed action as closely as possible and learning why that action was performed by extracting state-action pairs [58]. The technique has seen many applications for replicating human behaviour as features that make actions imperfect or appear human are also included. This allows agents that cooperate with an agent trained through imitation learning to be more robust to less-optimal strategies and behaviour[32].

Where this methodology falls short is in variability as the resulting agent will struggle to perform unique behaviours as it continues to perform learned actions repeatedly with little variance [58]. This has damaging effects on the performance of cooperative agents training with it as they can treat certain habits with greater importance than they have [32]. Potential approaches designed to address this include generating a larger pool of diverse behaviours by providing more informative training samples and through Generative Adversarial Networks (GANs) [84]. GANs work by including a discriminative classifier which attempts to identify whether a presented behaviour was generated or the ground truth in an iterative process. In addition, Imitation learning struggles to capture dynamic features such as how other agents may respond to an agent's behaviour and does not generalize well to new or noisy environments [86].

### 3.2.2 Inverse Reinforcement Learning (IRL)

Inverse Reinforcement Learning (IRL) attempts to imitate an observed behaviour by extracting a hidden reward function from it [7]. This can potentially result in very close imitations that generalize to new contexts if a suitable reward function is found, and remove the need for one to be manually specified. However, finding this reward function can be challenging and requires a fine balance between training with less data so that the model can better adapt to new situations and more data to achieve a better approximation of the reward function. In addition, the goal of the model can be ambiguous if the reward function of observed behaviour is difficult to specify or is dynamic [79]. Because of this, as well as problems with captures of ideal behaviours being often noisy, manual intervention and adjustments are often required to achieve a desired behaviour [118].

An extension of IRL known as Cooperative Inverse Reinforcement Learning (CIRL) aims to enhance the process of finding the hidden reward function by demanding greater involvement from humans and having them interact with the AI agent [79]. This involves the human teaching and correcting the agent during the training process so that the agent can better maximize the human's reward function. However, CIRL remains quite immature, and much of its value is unclear as it makes numerous assumptions about how the human player will act, such as how they are rational and will aim to convey information to the AI as much as possible [31].

### 3.2.3 Interactive Learning

Interactive learning is a novel method, that includes a human in the training process to help improve learning rates by taking advantage of prior knowledge, and to provide them with greater agency to shape the AI agent's behaviour to better meet expectations and goals [63, 102]. A useful side effect of this is improved sampling efficiency as less data is required to achieve higher levels of performance[8]. Human feedback can come in many forms of quantitative critiques which can be binary, or scalar values, to provide positive, or negative feedback for a specific action chosen by the AI [102]. Feedback can also be qualitative in the form of queries where humans choose sub-rewards for the model out of a set of proposed options, action advice where the user illustrates an optimal action and guidance where humans describe what they believe to be the goal at any given time-step[63, 8].

Challenges of this interactive learning are that feedback can be non-optimal, strongly biased, and be influenced by numerous environmental and psychological factors such as fatigue and loss of motivation [102]. These factors can cause inconsistencies in feedback which are especially pronounced when the frequency of interventions required is high, negatively impacting the training process.

### 3.2.4 Believable AI

The field of Believable AI is a field in computer science that involves the study of techniques that make AI agents behave similarly to humans. This field of study was largely active before the rise of machine learning and as such, uses many techniques to improve the believability of traditional state-based agents in games [143], though in recent years, techniques involving machine learning such as behaviour cloning and reinforcement learning [46] have begun seeing use. Some approaches to achieving more believable behaviour include giving AI agents a personality and role which biase their behaviour towards certain actions [159, 144, 21, 114, 154, 164], adding probabilistic behaviours such as jumping, rotating and movement [135] and copying behaviour from observed human players [123, 136]. In conjunction with techniques in making AI agents act more human-like, there must exist some measure of believability. A test that usually comes to mind is the 'Turing test' which is a test that measures how human-like an AI agent is. However, this test is very challenging to pass and remains a grand challenge for many AI researchers interested in producing AI with true intelligence [108, 82]. As such, many smaller tests were developed to test specific aspects of AI such as believability. One such test involves the use of a first-person shooter game 'Quake' where AI agents and humans play with one another across multiple levels while observed by human judges [108]. Judges will then rate how human-like each player was, with AI agents with a higher score being deemed more believable. Typical actions that resulted in AI agents receiving a lower score include having reactions that were too fast or being too accurate with their aim [108, 82].

## 3.3 Adaptating to Intent

Though the inclusion of a human proxy during the training process typically makes AI agents more robust in cooperating with humans in real time, humans can act very differently from one another. This can be partially resolved through diversity in human models[32] however agents trained with vastly different strategies struggle to effectively learn to cooperate with any and or have poor training results [176]. This heading will begin by exploring different approaches for measuring and identifying human intent, before describing meta-learning techniques that allow AI agents to adapt accordingly.

### 3.3.1 Measuring Intent

Approaches to measuring human intent can be found in the field of robotics where Mygography, the measurement of muscle activations, is commonly used to detect movement from specific regions of the body with the assistance of wearable sensors [109, 65]. An example of this is a power-assist glove that uses sensors to detect flexion angles of the user's joints to determine whether it should assist them [95]. Visual sensors can also be used to detect human gestures from which intention could be extracted from [128]. Some approaches to achieve this include model-based approaches where a kinematic model is placed into the scene to simulate human gestures, as well as heuristic-based methods, that use depth sensors to find a human profile and then discern the different components of the body to establish information about gesture. Many other approaches to measuring intent exist, however they share a common trait of attempting to paramaterise some factor that could infer intentionality. Once indicators of intent have been measured, their meaning can then be assessed.

### 3.3.2 Discerning Intent

One approach to discerning human intent is the use of Hidden Markov Models (HMM) which extracts information from observable states in the environment to infer the meaning of hidden states [115]. Besides the hidden nature of some states, the model assumes states abide by the Markov property like traditional

Markov models which means that the current state of a process is sufficient to predict potential future states without needing to draw upon previous states. HMMs have seen use in experiments involving service robots to classify gestures demonstrated by humans and found success even in context-sensitive and combinatorial actions such as pointing [128]. This method however struggles as state spaces grow larger and so usually has to be limited in some way such as constraining the number of gestural patterns or classes. A method addressing this uses hierarchical trees to represent human intentions and then infers a likely action through Bayesian inference to infer likely actions they will perform and reverse engineer their strategy [85].

### 3.3.3 Shared Mental Models

Shared Mental Models (SMMs) proposes that for humans and AI agents to co-operate, they must both have an understanding of their individual and shared goals [6]. Mental models are context-specific and/or context-sensitive often requiring specialized architecture which limits more widespread adoption and testing. Another limitation is the development of explainable AI which limits what conclusions can be drawn from these frameworks. These methods remain relatively unexplored with only a few minor experiments in the past decade [75, 80, 188], largely due to how abstract they are, with minimal concrete benefits. More developed approaches to model and predict human intentions and behaviours include Dynamic Function Allocation and Adaptive Automoation [29, 137, 101, 93, 16, 100, 94]. Though these methods have found success in areas such as aerospace [29] and navigating vehicles [100], they often require bespoke solutions for different contexts, struggle to generalize to different partners and only consider unidirectional adaptation with the AI adapting to the human.

### 3.3.4 Meta-Learning

Meta-learning is a set of diverse techniques that are concerned with designing robust algorithms that can learn and adapt to new environments more effectively [124]. Model-agnostic meta-learning (MAML) is a common approach that achieves this by calculating a set of parameters that is widely suitable for a variety of tasks so that when deviations occur, the model is equipped to learn quickly from a smaller number of new data through fine adjustments of weights and adapt well to new tasks [124, 70]. Some approaches develop new model architectures inspired by nature to improve learning such as Backpropamine which takes inspiration from how animals learn to improve adaptation to new settings in the form of neuromodulated plasticity [126], and other evolutionary strategies which accelerate policy convergence [148]. Other contributions to the field of meta-learning include Few-Shot learning approaches which optimize the ability to learn from limited samples [192], unique Hebbian Softmax models which improve the speed of neural networks by retaining important information through efficient memory structures known as 'neural caches' [140] and ensemble learning models which assist with generalizing models to improve performance in new scenarios [156].

## 3.4 Archetypal Analysis Variations

### 3.4.1 Deep Archetypal Analysis

Traditional linear AA models have been found to have limitations preventing them from optimally approximating certain datasets such as those where no prior knowledge exists. This is due to the limitations of linear separations of features, especially those of higher dimensions, in addition to having a strong, prior understanding of how many meaningful archetypes exist. A solution proposed to address this problem is the use of an encoder-decoder framework that learns and compresses latent information about the dataset [97]. This compression handles side information which provides insights about a dataset to guide the appropriate number of archetypes and dimensions. A property of this technique coined 'Deep Archetypal Analysis', is that it is generative, allowing for interpolations between archetypes to create new archetypal mixtures.

### 3.4.2 Probabilistic Archetypal Analysis

AA models are limited in their applications to datasets where observations exist in vector space and are real. Though this has a wide range of applications, it fails to handle certain data types such as binary data and probability vectors. An approach to address this is to loosen the restrictions employed by the original algorithm, allowing archetypal compositions to not perfectly recreate the original archetypes but be a sparse set that could explain the observations of a dataset [152]. As a result, the new algorithms allow of sparser archetypal compositions so that reconstruction can deviate slightly from archetypes, while also supporting different data types such as Bernoulli, Poisson and Multinomial distributed data. In doing so, additional observation types are supported, including Bernoulli, Poisson, and Multinomial observations.

### 3.4.3 Fast Archetypal Analysis

AA is generally limited by the speed of the algorithm. Some approaches include narrowing down the set of potential archetypal points or by preselecting them. This avoids expensive optimization considering all data points and limits them to a smaller set where a convex hull could be found [23]. This is possible as a large number of data points do not make good archetypal points. though in higher dimensions this becomes less effective as points start approaching the hull, it is always more efficient than the original algorithm. [38]. There is however a chance for a loss of precisions since only a subset of data points are considered for archetype selection.

## 3.5 Research Discussion and Open Problems

This section briefly explains these lacking areas and provides a launching point for new investigations. Having covered a wide field of topics, low-level improvements to current methodologies will not be covered and explanations will remain high-level.

### 3.5.1 Communication between AI and Humans

In the field of Cooperative AI, the goal is not necessarily to have the AI fulfill any specific task as specified by a human but to be capable of making decisions that could go against human expectations yet also serve a unified goal [34]. This Leader/follower behaviour either has the human wield complete control over the AI or a stubborn agent who commits to certain actions with or without the human [32]. Little research into developing a balance between the two roles has been done as it can prove quite challenging. This is largely due to the limited ways that humans and AI can communicate with one another, which limits the ability to negotiate and adapt to one another [103]. However, recent advancements in natural language processing pose great potential for the area of Cooperative AI, especially for the ability of AI to express their intentions to humans, something that has largely remained uni-directional [68]. This would enable more common use of social frameworks to enhance cooperation such as negotiation between agents [30], development in trust models of others [42, 10] and voting in decision making [44]. In addition, advancements in the field of 'Understandable AI' could improve the ability for AI and humans to cooperate by giving us a better indication of the effects of certain parameters and why certain decisions were made by the AI agent [49]

### 3.5.2 Diversification and Adaptation of AI behaviour

Cooperative AI research relies on the availability of models representative of human behaviour, whether synthetically developed or involving human involvement, to perform optimally. However, synthetic models face challenges in being able to generalize to new contexts effectively, accurately capturing realistic reward functions, and can be data inefficient [144]. Human involvement similarly faces challenges such as being time-consuming and potentially suffering from biases and inconsistencies. As such, there exists the need to advance both approaches to improve cooperative models. On synthetic models, techniques to develop diverse AI behaviours and AI agents that act similar to humans is an open problem with many techniques being explored to improve them such as by giving AI agents a personality and role which biase their behaviour towards certain actions [159, 144, 21, 114, 154], adding probabilistic behaviours such as jumping, rotating and movement [135] and copying behaviour from observed human players [123, 136]. Meta-learning techniques also have the potential to be used to improve the adaptability of AI agents but have yet to see much exploration in the domain of cooperative AI research [124, 70]. On involving humans in the training process, there have been a variety of approaches to improving the process by making it more efficient, by reducing the amount of involvement required to get accurate behaviour capture [102, 8], and making models more robust [58, 86].

### 3.5.3 Modelling Human Personalities and Intent

For AI models to improve in their ability to cooperate with humans, there is a need for them to be able to understand the intentions of humans such as through mental models [6, 75, 80, 188]. Much research has been conducted in measuring and discerning human intentions in robotics, however, the techniques have yet to be applied to cooperative AI or RL settings in general [115, 128, 85]. Existing methods in robotics for collecting information on human intent generally involve the use of wearable technology [109, 65] or sensors [128] which can enable the collection of more data suitable for informing the actions of AI in cooperative contexts. Another promising approach to developing better mental models of humans is through the use of clustering techniques such as Archetypal Analysis which has already seen applications in contexts outside of cooperative AI [47, 168, 66]. This allows human partners to be represented as combinations of archetypal behaviours which can be effective in informing AI models on how to best adapt to them.

## 3.6 Summary

As evident in the above subsection, many techniques have been explored for improving understanding human personalities, and adapting technologies to better meet the needs of humans. However, little of this work has been applied in the field of cooperative AI for improving the cooperativeness of RL agents which has largely focused on techniques to improve robustness, such as training with human proxies or a large volume of other AI agents, or focused on AI-AI teams. Though these are important contributing factors to effective cooperative models, more work needs to be done for better understanding and adapting to human partners.

This gap in research drives our work in this thesis which looks towards the use of clustering algorithms to develop better mental models of human partners in RL agents, to improve cooperation. Specifically, we look to the use of AA which amongst other clustering techniques, best matches theoretical understandings of human personalities. Using the mental models developed by this approach, we intend to improve the ability for RL agents to adapt to novel human partners and reduce the need for explicit communication through better implicit understanding of their partner's needs.

# Chapter 4 Overcooked AI: Implementation

This section outlines our proposed framework for a cooperative model that integrates archetypal analysis (AA) into a reinforcement learning agent. This section largely addressed (RQ1 - A, B). For assistance in visualizing the process, please refer to Figure 5.

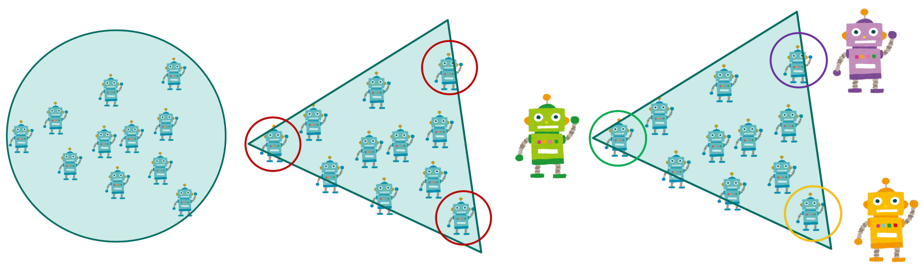

Figure 5: The framework of our AA agent. In the **left-most image**, we use fictitious co-play to train models and have the models complete episodes in the environment while tracking play data. Illustrated in the **centre image**, we then performed AA on the dataset to generate archetypes representing outlier playstyles. We then train separate PPO models to cooperate with each archetype as shown in the **right-most image**.

## 4.1 Environment

In reinforcement learning research, games are often used as testbeds for training and testing models due to their ability to manipulate environmental information. Though they do not always translate well to real-life environments, they are sufficient to serve as a benchmark to use for comparisons to other models. To assess the objectives listed, the environment for this experiment needs to adequately accommodate a variety of potential play styles and strategies, as well as allow for meaningful cooperation between agents. This typically means that the game should have a longer time horizon than that used in previous experiments in addition to a variety of meaningful objectives.

An environment that has seen increasing use for developing cooperative AI models is ‘Overcooked’, a multiagent environment that challenges multiple agents in their ability to cooperate [32, 162]. In this game, players act as chefs responsible for delivering as many dishes to customers within a given time frame. To do so, they must coordinate with one another in the kitchen by delegating tasks, sharing ingredients, and avoiding running into one another. A summary of this gameplay can be found in Figure 7, as well as the state-action space in Table 1. There are a variety of levels in the game with their unique layouts but a few objects that are common among all of them are as follows:

- Onion dispenser: A location in a level that when interacted with by a player, provides them with a single instance of an onion. If the player already has an item in their possession, then nothing will happen unless it is an onion in which case it will disappear.
- Plate dispenser: A location in a level that when interacted with by a player, functions similarly to the onion dispenser but instead provides plates.


Figure 6: The ‘Cramped Room’ layout we built upon, and conducted experiments with, from the ‘Overcooked ai’ environment developed by Carroll et al. [32].

- Tables: The most common location in levels. They can be interacted with by players to drop items such as onions and plates. This is generally done as to position items in a more convenient position for other players to interact with.

- Pot: A location in a level where players deposit onions. Once the pot has been filled with 3 onions, it will stop accepting them and begin to cook them. After a short duration, the onions will be cooked and the pot will enter a 'filled' state. While in this state, players need to have a plate to interact with it in which case, the player they have in possession will be transformed into a 'dish' and the pot will return to its default state.

- Counter: A location in a level where players deliver dishes to gain points.

Table 1: State-Action Space for Overcooked

| **State Space** | **Description** |
|---|---|
| Player Positions | Locations of all player characters on the game grid. |
| Ingredient Positions | Locations of all ingredients (on table, or being carried). |
| Dish Statuses | Status of dishes (raw, cooked, plated, served). |
| Kitchen Layout | Configuration of the kitchen (tables, cooking pots, plate dispenser, onion dispenser, counter). |
| Timers | Relevant countdown timers (cooking timers). |

| **Action Space** | **Description** |
|---|---|
| Movement | Actions to move in four directions (up, down, left, right). |
| Interaction | Actions to interact with objects (picking up, putting down, cooking, serving). |
| Use Items | Actions to use or manipulate items (placing ingredients in pots, taking cooked food out of pots, plating dishes). |

Overcooked serves as a great testbed for our purposes as it has a high level of complexity, with multiple tasks to be performed simultaneously and others that may force coordination between agents. This complexity enables the emergence

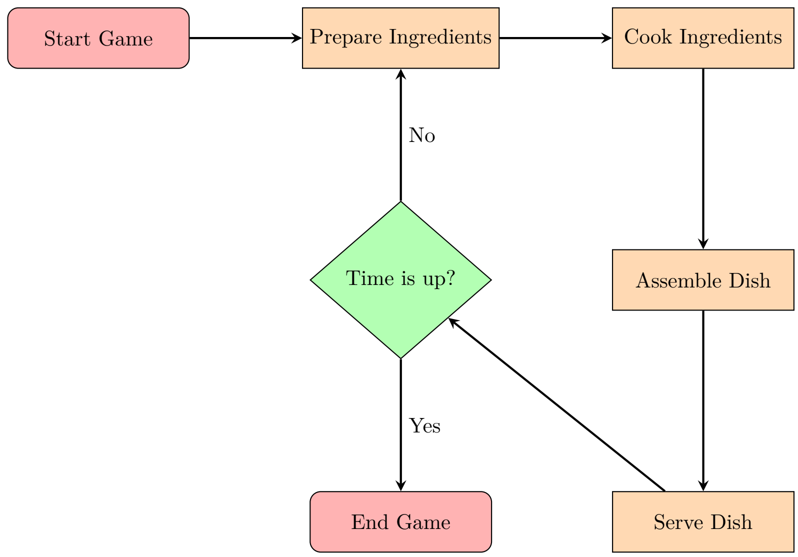


Figure 7: Gameplay outline of Overcooked showing the core actions that can be done in the game as well as the order in which they are done.

of more diverse play patterns, which in turn leads to more meaningful archetypes when applying archetypal analysis. In addition, the multi-agent nature of the environment is ideal for assessing what archetypes may perform better with others, which can be invaluable when attempting to coordinate with an unseen player. This is supported by the sparse nature of rewards that encourages agents to work together to achieve long-term goals as opposed to acquiring fast rewards they could get themselves.

We used the environment implementation developed by Carroll et. al. [32], as seen in Fig. 6, with a few adjustments to facilitate our custom AA agent. This included a few data structures to hold information on each episode as well as information on the archetypal profile of each player regardless of whether they were human or AI. These were crucial for the implementation of our AA agent which is covered in the following section.

## 4.2 Archetypal Analysis Agent Implementation

Our approach in integrating AA to develop cooperative AI models is by using it as a heuristic in an ensemble framework, where it can be used to determine the most appropriate model action to use. To achieve this, we first created a dataset of playthroughs by representative human players to run AA on and find archetypal playstyles. Once this was developed, we then trained an RL model with each archetypal playstyle so that they were optimized to cooperate with them. This would leave us with N number of RL models where N is the number of archetypal playstyles which we would shift between to select actions during runtime using ensemble learning.

### 4.2.1 Dataset Synthesis

To produce the initial dataset of playstyles, we chose to create RL models that were representative of human playstyles as opposed to recording real human playthroughs. This was because it would be expensive and time-consuming to have a human play through a large number of levels to create the dataset. To create the RL models, we took inspiration from Strouse et al. [162] and their use of fictitious co-play models and took a similar approach in creating this initial dataset to avoid the use of human data. This involved training 5 self-play agents using a Proximal Policy Optimization policy (PPO) that is checkpointed during the process to represent levels of player skill. These models then play with one another and the data from each playthrough is saved. Through observations collected during recordings of multiple playthroughs of the Overcooked environment by human players, we found that the most relevant features conducive to determining the archetype of a player were:

- Number of objects placed
- Number of objects boiled

- Number of soup delivered
- Number of soup plated

Each of the 5 PPO agents was initiated with a random seed and starting position, then trained for 10,000 timesteps and checkpointed after timesteps 2500, 5000, and 7500, producing 20 different models, converging at largely different policies. These models would then play a random number of playthroughs with themselves in an Overcooked level ranging from 100-150 times each. In the end, we produced 2647 playthroughs which served as our dataset representing different approaches to the environment.

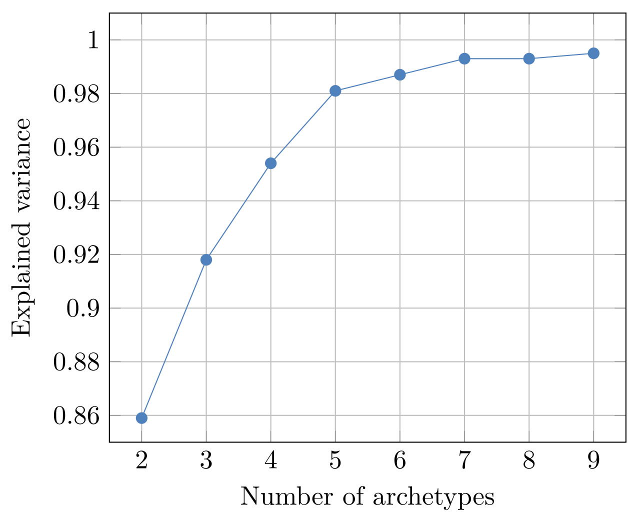


Figure 8: Comparison of different number of archetypes

### 4.2.2 AA on Synthetic Dataset

We then perform archetypal analysis on this dataset, which provides us with K number of archetypes, where K is an arbitrary integer we choose. To aid us in choosing an effective number of archetypes, we calculated the explained variance of different numbers of archetypes, which can be seen in Fig. 8. Evidently, the

Table 2: Archetype Profiles

| **Arch-type** | **Features** | | | |
|---|---|---|---|---|
| | ***objects placed*** | ***objects boiled*** | ***soup delivered*** | ***soup placed*** |
| A1 | 0.000000 | 0.770684 | 0.582924 | 1.000000 |
| A2 | 0.000000 | 0.846997 | 0.834530 | 0.000000 |
| A3 | 0.954288 | 0.000000 | 0.000000 | 0.295429 |

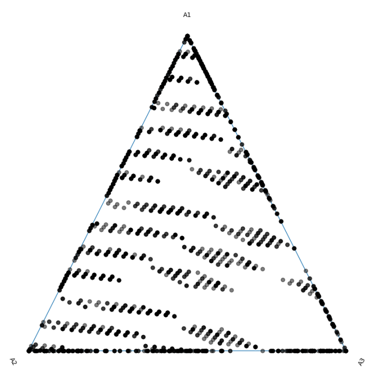


Figure 9: Datapoints expressed as a convex combination of 3 archetypes.

benefits of additional archetypes taper off after 5 archetypes with 3-5 archetypes explaining most of the variance in the dataset. Using fewer archetypes would fail to encode all the information from the dataset while using more archetypes would take away the benefits of the dimensionality reduction effects afforded by the algorithm. With the original feature set having a dimensionality of 4, we chose to proceed with 3 archetypes to take advantage of the dimensionality reduction benefits of archetypal analysis. The profiles of the archetypes generated with K = 3 archetypes are seen in Table 2 and the data points in the feature-set can be expressed as a convex combination of these 3 archetypes shown in Fig. 9.

#### 4.2.3 AA Agent Training

Afterward, we trained an RL model with a PPO policy to cooperate with the model that had the closest alignment to each archetype for 10,000 timesteps. The result of this was 3 policies that were each tailored to cooperate with one of the 3 player archetypes produced previously. To have the agent adapt to these player archetypes and archetypal mixtures, we take inspiration from Villareal et al. where ensemble learning was used to discern different styles of human drawings and inform which classification model their framework should use [176].

Ensemble learning is an approach typically used for improving the predictive performance of multiple classification models [156]. This involves training a variety of classification models and combining their predictions to produce a joint prediction as seen in Figure 10. This could be done through voting where the most common prediction is used or stochastically chosen from, through an average result of the predictions, or some heuristic. The result of this is that the combined model is most robust, mitigating overfitting and reducing bias that may occur when solely using the prediction of any single classification model.

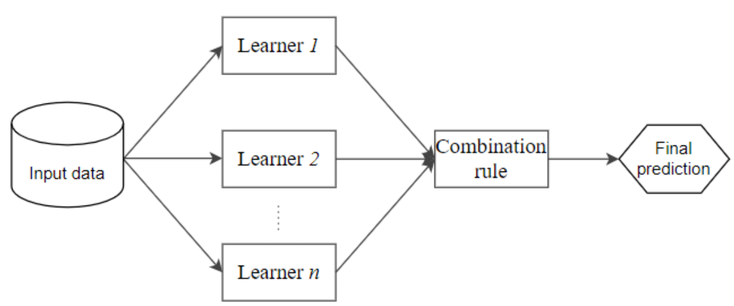


Figure 10: A visual representation of Ensemble learning (A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects)[163]

For determining our agents next move, we chose a stochastic approach to action selection. During runtime, our AA agent observes the player's alignment to one of the aforementioned archetypes through a least-squares algorithm comparing the player's current play data scaled by time. Using the resulting vector representing alignment, the agent stochastically selects the appropriate cooperative model action weighted by the player's proximity to the given archetypes. For example, based on a player's actions, performing a least-squares operation may result in a distribution of [0.001, 0.87, 0.129] representing the convex combination of 3 archetypes. The model will then select the 1st archetype with 0.001 probability, the 2nd archetype with 0.87 probability, and the 3rd archetype with 0.129 probability.

### 4.2.4 Other Design Considerations

One of the reasons we chose not to use human playthroughs for our initial dataset is that it is expensive and time-consuming. However, another reason for the decision was that we would need to train a model to cooperate with an archetypal playstyle afterwards. If we had human playthroughs, the player who was closest to an archetypal playstyle would then need to play thousands of games with the cooperative agent for training purposes. This intensive requirement for repeated play sessions with a human partner was deemed impractical and unacceptable for our project's scope and timeline. As such, we decided to produce a variety of human-like playstyles artificially using the Fictitious Co-play technique.

An alternative approach of producing artifical playstyles was the use of a smaller amount of human playthrough data to train RL models through behaviour cloning or inverse reinforcement leraning and then using the resulting model to produce playthrough data. Though this was considered, we ultimately decided against this approach for several reasons. Firstly, using Fictitious co-play allowed us to make more direct and meaningful comparisons with self-play models that do not rely on human data that we are benchmarking against. Secondly, though one of the benchmark models we aimed to compare against does utilize human data as mentioned in the following section, the comparison would be unfair as we do not have access to the specific human data used for its training. This lack of access is discussed more in the conclusion in the Limitations section of Chapter 8, and details about the benchmark models are provided in the following section.

To ensure consistency and fairness in our comparisons, we chose to use self-play agents with a PPO policy to generate our models. PPO is a widely used RL algorithm that balances exploration and exploitation effectively, making it a robust choice for training cooperative agents. Furthermore, it is also the policy that was used by the self-play benchmark model, and by using it, we aimed to isolate the impact of our framework and reduce confounding variables related to the underlying RL algorithms and training parameters.

## 4.3 Benchmarks

To assess the capabilities of our custom cooperative agent framework, we compare it with other models that have been used for cooperation. A variety of agents were used in the experiment including 2 benchmark models from Carroll et al. [32]:

- Self-play agent: This benchmark agent was trained from scratch with itself using PPO. It had access to information about its state, such as what object it was holding, as well as the state of the environment, such as the number of objects in the pot. To expedite training, it was given rewards for positive intermediate actions, including putting onions into the pot and picking up soup with a dish.
- Human-Trained agent: This benchmark agent was developed by first training a model to act as similarly to a human as possible using behaviour

cloning, a technique where the model learns a policy from demonstrations. The agent then trains with this model using PPO and implements a model-based planner that uses a hierarchical A* search to act and strategize optimally in response to the policy of their partner.

- Random action agent: This agent selects an action to perform at complete random. It is not expected that this agent will perform well and it largely acts as a point of comparison for participants to evaluate the cooperativity of a partner.
- AA agent: This ensemble agent was trained using our custom framework described previously under the Implementation section. It makes use of archetypal analysis to select the appropriate model action in response to their partner's perceived playstyle.

## 4.4 Overcooked AI: Metrics

This section outlines our research into techniques for evaluating AI alignment to human playstyles, mainly addressing research questions (RQ2 - A, B, C). It begins with a review of existing methods before exploring potential approaches from other research fields that could be applied comparing the alignment of RL models.

### 4.4.1 Quantitative Metrics

When comparing the performance of Cooperative RL models, researchers have employed various metrics and approaches to assess their effectiveness [121, 32, 162, 120, 191]. Carroll et al., in their experiment on the Overcooked testbed, utilized metrics such as cumulative reward and average rewards per episode to evaluate the performance of their AI models [32]. Other experiments ran on the same game such as those run by Strouse et al. instead used the number of dishes that were delivered before the time limit for the level ran out, which though provides a less detailed metric, is more understandable [162]. In summary, common methods used to assess agent cooperation and alignment with human goals generally involved comparisons of the score that they achieved, or the rewards the model received, to assess how well an AI agent was able to align themselves, and adapt to human player goals - the expectation, being, that those that scored higher were better able to align to a human player's personality.

### 4.4.2 Qualitative Metrics

Though these approaches see common usage and can be useful to compare the performance between RL agents, they overlook factors unique to cooperative contexts such as perceived cooperation. For instance, in human-robot interaction scenarios, the perception of cooperation from users can significantly influence their satisfaction and engagement with the system, regardless of individual agent performance metrics [173, 157, 172, 74, 9, 146]. In experiments run by Van den Bosch et al., interviews and questionnaires are used to evaluate the experience of humans who collaborated with their AI system, asking questions about morality and trust [173]. Ulfert et al., similarly use interviews to elicit the thoughts of participants about the AI systems they collaborated with as well as how they reached those conclusions [172]. The latter experiment also includes observations about their participants, noting habits and behaviour patterns to gain further insights about the cooperativity of their systems. Other previous experiments have also included assessing the intrinsic motivation of participants, by containing questions related to the past experiences of the participant [146]. These qualitative techniques provide rich data which allow for more analysis and conclusions to be made in comparison to quantitative approaches. Despite this, some approaches such as interviews and the collection of observational data have yet to see much use in the context of evaluating cooperative RL models.

### 4.4.3 Mixed Methods

Although the aforementioned qualitative techniques often provide deeper insight into generally how cooperative systems are, it can be difficult to compare the level of cooperativity between different AI systems in comparison to quantitative measures. An experiment that demonstrates a promising approach to remedy this issue is one run by Strouse et al. who use closed questions in combination with a 5-point Likert scale when asking participants to evaluate their preference for different AI partners [162]. Furthermore, past experiments also evaluate the alignment of AI agents to human expectations, by reviewing actions that were performed by the agent during a playthrough with a participant, and asking participants whether the action aligned with what they wanted the AI to do, and whether they believed the action was sufficiently cooperative [162]. This approach successfully takes advantage of the benefits of qualitative data while also providing a quantitative measure to compare different AI agents using, in this case, the number of reviewed actions that aligned with the player's expectations of the AI agent. Therefore, incorporating measures of perceived cooperation, such as user feedback or subjective evaluations of AI agents, alongside traditional, quantitative performance metrics can provide a more comprehensive understanding of the effectiveness of RL models in cooperative settings.

### 4.4.4 Experiment Implementation

We incorporate the above methods in our experiment, with the data collected during the experiment including observational notes on player behaviour and trends, scores they achieved with each AI, and a questionnaire that the participants filled out. The questionnaire was filled by players after each playthrough with questions varying from those assessing their intrinsic motivation, and evaluating the experiences of playing games, to open questions that assess the levels of confidence and trust that players have of their team partners. The questionnaire also asks participants to rate the cooperativity of each agent they partner with using a 5-point Likert scale, giving us a qualitative measure in which to compare the perceived level of cooperation for each agent. The questions for each agent included:

- How cooperative did you feel your partner was on a scale of 1-5?
- What factors contributed to you reaching the above conclusion?

General questions that were asked include:

- How experienced are you with playing cooperative video games?
- What differences stood out between playing with a human and an AI?
- Any final comments you would like to add?

## 4.5 Chapter Summary

In this section, we described the 'Overcooked' environment that we will use as a testbed for our experiments which will be explained in Chapter 5. We justified that this is an appropriate environment for training and evaluating cooperative RL agents due to facilitating a variety of strategies and multiple approaches of cooperation between players.

Following this, we provided a detailed explanation of how our AA agent was implemented through the process of creating a synthetic dataset of playthrough data, conducting AA on the dataset and then training multiple cooperative models that constitute the AA agent. We then outline other models we used in our experiment that acted as benchmarks to evaluate the performance of the AA agent on.

We continue with a discussion on metrics previously used to evaluate the performance of cooperative RL agents with human partners, and explain why they are often insufficient due to overlooking factors such as alignment in goals. Conluding this section is an outline of the combination of qualitative and quantitative measures we implemented in our experiment.

# Chapter 5 Overcooked AI: Experiment

Having described the setup and implementation of the experiment in the previous chapter, under this heading, we delve into the procedure of the experiment ran and the demographics of the participants involved.

In this experiment, we aimed to evaluate the effectiveness of our AA agent in the 'Overcooked' testbed, compared with the previously described benchmark models, answering the research questions outlined in the beginning of the thesis (RQ1 - C,D).This 'effectiveness' is evaluated using two measures: the score the model and its human partner achieved in a playthrough, and the cooperative rating that its human partner gave it afterwards. We also apply our research into techniques for evaluating AI alignment with human player personalities (RQ2 - D).

## 5.1 Procedure

The experiment was held in person at a university computer lab and administered in a supervised format. This allowed for experimental control and for observations by the experiment host in regards to actions users performed as well as to provide context for the largely quantitative data collected. Upon entry, participants were greeted and informed of the experiment's procedure. They would then be required to complete a written consent form and be given an initial safety brief. They were then directed to a computer with instructions for how to play the game while a host would answer any questions they had. To ensure that players understand the directions, they would play a test game that was not recorded with a random AI agent. They were allowed to play as many test games as necessary until they felt confident in playing the game. This was crucial to avoid as much bias as possible caused by familiarity with the game.

Once ready, the host would randomly pair the participant with an AI agent among the list of agents listed above, whom they would then play a game with. Each game lasted for 30 seconds, after which they were prompted to complete the relevant sections of the questionnaire and the experiment host would note down the score that was achieved and any observations they had. The experiment host may decide to ask questions about the playthrough after it was completed but no questions were asked during the playthroughs to avoid distracting the participant.

Upon completion, the host would randomly change the AI agent to one that had not already been chosen and then prompt the participant to begin again when they were prepared. This procedure would continue until the participant had been partnered with each AI agent, at which point they would be debriefed and the experiment was concluded. This followed a within-participant design, that is, each participant played with each style of agent, though the order in which they were paired were random. This allowed us to isolate biases to each individual, evaluate the scores they achieved and the unique preferences they had.

## 5.2 Participants

Participants were recruited through announcements distributed across university channels, including club communications, and on local game industry forums and groups. These recruitment messages were uniformly disseminated across public forums to maintain a non-personalized approach, thereby minimizing potential biases and ensuring voluntary participation.

Our goal was to acquire approximately 20 participants and then adjust this number based on data saturation analysis. That is, after each experiment, results were reviewed and a Bayesian Mann-Whitney U Test was conducted. If additional participants did not change the Bayes factor by at least 10% we would halt recruitment, otherwise initiate another wave of recruitment. For this experiment, our first wave of recruitment concluded with 16 participants and after the aforementioned data saturation analysis, we found it unnecessary to conduct further recruitment efforts. These tests can be found in the Results section.

The participants we acquired fell under the 20-30 year old range with 75% of participants between 20-25 and the remaining 25% between 25-30. Participants had a diverse range of experiences in playing cooperative games with 12.5% rating themselves as beginners to cooperative games, 56.25% as having intermediate experience, and 31.25% as having advanced experience. As a result, we saw a good variety of perspectives on cooperative behaviour and several different strategies to succeed in the game.

# Chapter 6 Overcooked AI: Experiment Results

In this chapter, we provide a summary of the results we achieved during the experiment. This includes quantitative comparisons between the scores each AI agent achieved as well as the cooperative rating that participants gave them. Qualitative data such as observations of participants and the responses they provided in the questionnaire will also be summarised.

## 6.1 Quantitative Results

The following section covers the quantitative results that were collected through participant playthroughs and the rating participants gave to agents in the questionnaire. We conducted Bayesian, non-parametric ANOVA tests on the two quantitative measures of agent performance, that being the scores they were able to achieve with human partners as well as the cooperativity rating participants gave each agent in the questionnaire. Though we also experimented with a standard non-parametric ANOVA test, we found the Bayesian equivalent had similar results but provided additional information as a result of the Bayes factor. We then performed paired sample tests using a Bayesian Wilcoxon signed-rank test to gather more specific data on comparisons between models. We first begin with the tests on the scores that each model was able to achieve and then continue to discuss the equivalent for the cooperative ratings each model received.

### 6.1.1 Score

From the non-parametric ANOVA test shown in Table 3 and visualized in Figure 11, the BF (Bayes Factor) when comparing AA to Human-Trained, AA to Self-play and Self-play to Human-Trained agents were all equivocal, with a BF10

Table 3: Post Hoc Comparisons - Agent Type

| | | Prior Odds | Posterior Odds | $BF_{10,U}$ | error % |
|---|---|---|---|---|---|
| AA | SelfPlay | 0.414 | 0.122 | 0.294 | 0.013 |
| | Random | 0.414 | 14.910 | 35.997 | $7.802 \times 10^{-7}$ |
| | Human_Trained | 0.414 | 0.640 | 1.545 | $3.122 \times 10^{-6}$ |
| SelfPlay | Random | 0.414 | 3592.837 | 8673.877 | $2.548 \times 10^{-7}$ |
| | Human_Trained | 0.414 | 0.575 | 1.387 | $3.761 \times 10^{-6}$ |
| Random | Human_Trained | 0.414 | 5113.135 | 12344.200 | $1.359 \times 10^{-7}$ |

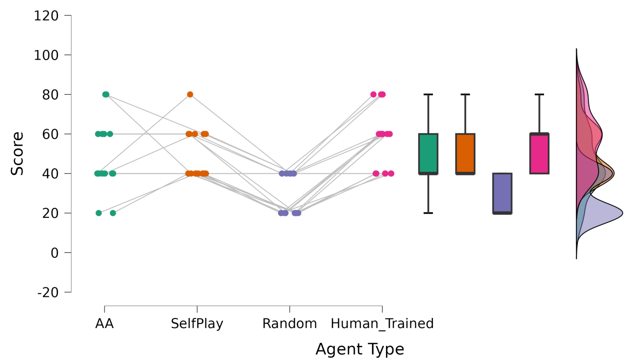


Figure 11: Raincloud plot of agent scores

$< 3$ representing insufficient evidence to prove that they are not equal. This suggests that there are no significant differences in their performance given the sample size. Following the same measure for significance, we found that there was sufficient evidence that random-action agents were significantly different from the other 3 agents.

Looking at the distribution of results in Table 4, we found there was substantial variability in the score of the AA agent, which had a standard deviation of 17.7, as well as the human-trained agent with a standard deviation of 15. This is in comparison to the random and self-play agents which had noticeably lower standard deviations.

We then conducted a Bayesian Paired Samples T-Test as seen in Table 5 to find additional details between the agents. Though this did not produce any new

Table 4: Descriptives of Score Data

| | | | | | | 95% Credible Interval | |
|---|---|---|---|---|---|---|---|
| Agent Type | N | Mean | SD | SE | CoV | Lower | Upper |
| AA | 16 | 47.500 | 17.701 | 4.425 | 0.373 | 38.068 | 56.932 |
| SelfPlay | 16 | 50.000 | 12.649 | 3.162 | 0.253 | 43.260 | 56.740 |
| Random | 16 | 28.750 | 10.247 | 2.562 | 0.356 | 23.290 | 34.210 |
| Human_Trained | 16 | 56.250 | 15.000 | 3.750 | 0.267 | 48.257 | 64.243 |

results, it did reinforce the findings we previously found with the ANOVA test. Pair-wise comparisons of the different agents can be found in Figure 13. We did not include the comparisons between self-play and human-trained models with random agents as they were not the focus of the experiment.

Table 5: Bayesian Wilcoxon Signed-Rank Test Score

| Measure 1 | | Measure 2 | $BF_{10}$ | W | Rhat |
|---|---|---|---|---|---|
| AA | - | SelfPlay | 0.527 | 8.000 | 1.000 |
| | - | Random | 37.910 | 55.000 | 1.002 |
| | - | Human_Trained | 1.898 | 10.000 | 1.000 |
| SelfPlay | - | Random | 374.649 | 105.000 | 1.018 |
| | - | Human_Trained | 1.421 | 4.000 | 1.000 |
| Random | - | Human_Trained | 377.061 | 0.000 | 1.012 |

### 6.1.2 Cooperative Rating

From the non-parametric ANOVA test shown in Table 6 and visualized in Figure 12, we found that when it came to ratings of cooperativity, the AA and human-trained models were equivalent with a BF10 $< 3$. Using the same measure, the self-play model performed significantly worse than AA and human-trained models but better than the random action model which performed overwhelmingly poorly in comparison to the other models.

Table 6: Post Hoc Comparisons - Agent Type Rating

| | | Prior Odds | Posterior Odds | $BF_{10,U}$ | error % |
|---|---|---|---|---|---|
| AA | SelfPlay | 0.414 | 2.643 | 6.380 | $8.686 \times 10^{-7}$ |
| | Random | 0.414 | 1675.453 | 4044.901 | $3.079 \times 10^{-7}$ |
| | Human_Trained | 0.414 | 0.202 | 0.487 | 0.019 |
| SelfPlay | Random | 0.414 | 5.042 | 12.172 | $7.208 \times 10^{-7}$ |
| | Human_Trained | 0.414 | 106.048 | 256.021 | $2.228 \times 10^{-7}$ |
| Random | Human_Trained | 0.414 | 4086.442 | 9865.545 | $2.155 \times 10^{-7}$ |

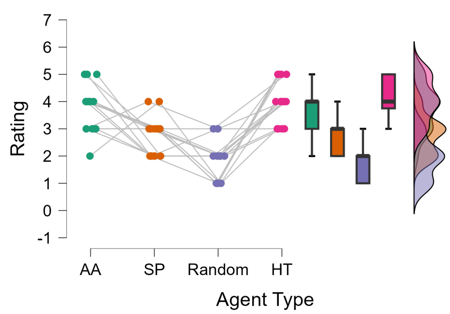


Figure 12: Raincloud plot of agent scores

Analysing the details of the cooperative rating results in Table 7, we found that similarly to the score results, the AA agent had the highest variance in their ratings, though they were closer in line with the variance of the other agents this time.

We then conducted a Bayesian Paired Samples T-Test as seen in Table 8 to find additional details in the pair-wise comparisons of the cooperative ratings

Table 7: Descriptives of Rating Data

| | | | | | | 95% Credible Interval | |
|---|---|---|---|---|---|---|---|
| Agent Type | N | Mean | SD | SE | CoV | Lower | Upper |
| AA | 16 | 3.750 | 0.856 | 0.214 | 0.228 | 3.294 | 4.206 |
| SelfPlay | 16 | 2.750 | 0.683 | 0.171 | 0.248 | 2.386 | 3.114 |
| Random | 16 | 1.875 | 0.719 | 0.180 | 0.383 | 1.492 | 2.258 |
| Human_Trained | 16 | 4.063 | 0.772 | 0.193 | 0.190 | 3.651 | 4.474 |

received by different agents, which can be found in Figure 14. We did not include the comparisons between self-play and human-trained models with random agents as they were not the focus of the experiment.

Table 8: Bayesian Wilcoxon Signed-Rank Test Rating

| Measure 1 | | Measure 2 | $BF_{10}$ | W | Rhat |
|---|---|---|---|---|---|
| AA | - | SelfPlay | 7.729 | 90.000 | 1.004 |
| | - | Random | 1450.546 | 120.000 | 1.007 |
| | - | Human_Trained | 0.521 | 20.000 | 1.000 |
| SelfPlay | - | Random | 16.759 | 62.000 | 1.006 |
| | - | Human_Trained | 60.059 | 0.000 | 1.002 |
| Random | - | Human_Trained | 173.140 | 0.000 | 1.021 |

In summary, the AA agent performed similarly to self-play and human-trained models regarding the score they were able to achieve during the experiment. Additionally, the AA agent achieved similar results to the human-trained model in cooperative ratings it received but was significantly better than self-play and random action models. In both score and cooperative rating measures, the AA agent had the greatest variance in its results.

## 6.2 Qualitative Results

The following set of results consists of the qualitative components of the experiment, acquired through observations by the experiment host as well as responses provided by participants in the questionnaire. In this section, we discuss the feedback provided and observations made for specific agents. Sample responses from the questionnaire shown in this section were chosen based on the relevancy of the feedback and to represent novel points made. That is, we chose not to include feedback commenting on factors outside of the cooperativity of the agents, those that were phrased poorly, or redundant points.

### 6.2.1 AA Agent

The feedback we received for the AA agent summarised in Table 9 was largely positive, with many comments recognizing that the agent made distinct attempts to cooperate with its partner's strategy. A few comments also mentioned that the AA agent would change its strategy multiple times throughout a playthrough to best match the strategy of the participant. Through observation, we found that this experience generally occurred to users who had more experience with cooperative games as they attempted to test the limits of the AA agent's ability to cooperate with them. As such, we believe that in response to (RQ1 - A) that AA is effective for classifying the personalities and overarching strategies of players.

The general theme of negative feedback regarding the AA agent was that it would sometimes suffer from periods of indecision. This often took the form of performing an action conducive to a strategy and then proceeding to perform an action towards another strategy which was frequently counter-productive such as placing an ingredient near the pot for the player's convenience before picking it up again and putting it into the pot themselves. Through observations, we found that this generally occurred to players whose archetypal mixture was close to an even distribution of each archetype, as due to the stochastic ensemble approach to action selection, the likelihood of multiple actions being conducive to a single strategy is quite low.

Table 9: AA Agent Feedback Quotes

| Data Source | Participant Quotes |
|---|---|
| Questionnaire | "The AI moved randomly around the workspace, blocking my character multiple times. It picked up items and used them correctly but sometimes it would put them down and do nothing." |
| | "AI does wait for me to do my role, And sometimes I put onions and it waits for me. Overall feels like its trying to work together. Except sometimes it goes everywhere and blocks my way which is counter-productive." |
| | "It did actions in sync with me at each stage of the game but we struggled at points with stuff needing to be done but nobody holding the right thing." |
| | "We swapped roles quite a bit but it worked well." |
| | "Moved slower and was easier to work with. Predictable and felt more cooperative." |

### 6.2.2 Human-trained Agent

The Human-trained agent received similarly positive feedback in general as seen in Table 10, with a majority of comments by participants praising its ability to adapt to their strategy. For most partners, the Human-Trained agent would be able to continually adapt to changes in strategy though sometimes, they would change their strategy and take over the player's role.

The main complaints by participants with the Human-trained agent were that it would often block them from performing actions during playthroughs. Through observation, this would often occur when the player and the AI agent need to pass each other to get to a given object, with the Human-trained agent frequently taking on a more assertive personality, not backing down on its current trajectory and forcing the player to change paths should they wish to avoid a stalemate where both of them remain stationary.

Table 10: Human-Trained Agent Feedback Quotes

| Data Source | Participant Quotes |
|---|---|
| Questionnaire | "The AI seemed to pick up onions when I went to collect dishes and vice versa which helped a lot to complete the level. It was also very quick. However there were moments where it blocked my character from moving." |
| | "I often find that my partner blocks my way and usually doesn't move out of the way. But I do notice that it considers what I am currently holding so some collab there but I can't communicate to it." |
| | "We took turns doing tasks and the AI knew what object o have based on what I had and what went in the pot." |
| | "We had a good system, me on dishes, him on onions, but he changed roles at the end which slowed things down." |
| | "Helpful because he placed the plates in a convenient location." |

#### 6.2.3 Self-play Agent

The Self-play agent generally received negative feedback from participants regarding its ability to cooperate with them as partners. One major theme in the responses found in Table 11 were that the Self-play agent tried to do everything themselves, completely disregarding the player, forcing them to try and adapt. This is expected as these agents only train with an idealized partner and thus expect their partner to perform the same actions that they have deemed optimal. Through observation, they act quite stubbornly and scripted, which leads to lower cooperativity ratings but still perform quite well in regards to the score they achieve as it forces their partner to adapt to them.

This would frustrate many of the participants who would attempt to assist the AI initially but lose motivation to once they realized that they were being ignored. These attempts often involved placing ingredients at convenient positions for the AI to use which would be ignored with the AI continually grabbing ingredients from the ingredient dispenser despite being further away. Furthermore, another observation that we made was that collisions between participants and the Self-play agent were very frequent and would also frustrate participants, demotivating them from playing further.

Table 11: Self-play Agent Feedback Quotes

| Data Source | Participant Quotes |
|---|---|
| Questionnaire | "The AI moved randomly around the workspace, blocking my character multiple times from moving. It also picked up items and didn't use them which confused me. The AI was quick, which helped occasionally." |
| | "It doesn't feel like AI wants to work with me compared to the last AI [Human-trained]. I tried putting ingredients with them and they didn't even wait for me. It feels like the AI wants to do that role only so I adapted. " |
| | "The bot did everything and didn't give me the chance to place food in the pot." |
| | "Similar issues to the last one [Random-action] but was more flexible with their role." |
| | "Sometimes/often got in the way, blocked me when i tried to finish/help with the soup. Just running back and forth - annoying." |

### 6.2.4 Random-action Agent

The Random-action agent expectedly received poor feedback from participants due to not performing many actions conducive to the success of the team. The main points of feedback as summarised in Table 12 were that the Random-action agent did not know what to do and just simply walked around cluelessly. A few responses instead mention that the AI took a passive approach to co-operation, expecting the participant to initiate. This trend was also observed as participants would mention it verbally in passing through comments such as "I'm guessing this is the passive one".

Through observation, we found that the Random-action agent would occasionally receive relatively high cooperative ratings due to its tendency to avoid collisions with the player. These ratings would generally come from participants who were less experienced in playing cooperative games as they would be largely fixated on their actions rather than being concerned with the actions of the AI agent.

Table 12: Random-Action Agent Feedback Quotes

| Data Source | Participant Quotes |
|---|---|
| Questionnaire | "AI felt dumb, blocking my path and actions. At first easy to understand/avoid cos they stayed in a certain section, but then after it felt like they "didn't know what to do" - stopped me from delivering soup." |
| | "The AI moved randomly around the workspace, blocking my character multiple times. The AI would pick up items and do nothing with them sometimes." |
| | "The bot felt like it was expecting me to do all the tasks. It didn't provide big contributions to the pot." |
| | "He was good at the dishes but he blocked the pot for long periods of time." |
| | "Annoying and walking around not doing anything." |

## 6.3 Results Summary

The qualitative trends found through participant playthroughs of the experiment and responses from the questionnaire are summarised as follows:

1. In regards to the score achieved, a Bayesian, non-parametric ANOVA found that there is no significant evidence between the performance of the self-play, human-trained and AA agents. There was however greater variation in the score of the AA agent in comparison with the benchmark models, as evident by a higher standard deviation.

2. In regards to the cooperating rating each agent received, a Bayesian, non-parametric ANOVA found that the AA agent received significantly higher ratings compared to the self-play and random agents, and similar cooperative ratings to the human-trained agent.

To summarise overarching qualitative trends in the questionnaire responses and the observations made during the experiment by experiment hosts:

1. AI obstruction of player actions had a significant effect on the cooperativity ratings an agent was given and would greatly frustrate participants. This would often overshadow otherwise great performance and good decision-making processes by the AI agent as there were situations where the human-trained agent or AA models that performed well in score and cooperation received poorer ratings in cooperativity if they obstructed player movement or actions.
2. Participants with more experience playing cooperative games would pay greater attention to the actions of their AI partner, while in comparison, those with less experience would largely focus on their actions. This would result in occasionally higher than-expected opinions by less experienced players for poor-performing agents such as the Random-action agent as well as agents that appear competent such as the Self-play agent.
3. The AA and Human-trained agents were able to significantly better adapt to their partner's strategies compared to the Random-action and Self-play agents. The Human-trained agent was typically more robust and smooth in its adaptations to the player while the AA agent made more distinct shifts in strategy in comparison at the cost of more instability.

# Chapter 7 Discussion

This section will examine the findings of the results section and discuss their potential implications. This will be done considering the experiment structure, explaining how it affected the results and their significance.

## 7.1 Measuring Effective Cooperation

Effective cooperation in our experiment is the combination of how well an AI agent can serve an unspecified, unified goal with a human partner, and how cooperative they are perceived by the human partner. Thus, measuring the score that was achieved by an AI agent is not sufficient in encapsulating how capable it is in cooperating with a human partner, as a high score could have been achieved without collaboration between them. This is evident by the higher cooperative rating the AA agent received compared to the self-play model, despite both having similar scores. This is especially the case for the overcooked environment we used as it is an environment where cooperation is optional, that is, it is possible to achieve a reasonable score as an individual. Scores, however, may be a better indication of the cooperativity of an agent in environments where cooperation between partners is mandatory such as certain overcooked levels where scoring is impossible without mutual collaboration.

### 7.1.1 Measuring Cooperative Perception

Some qualitative measures used in our experiment include questionnaires and observational data which effectively capture the complexity of what makes up cooperation. The questionnaires were able to capture participant thoughts on the agents they collaborated with, gaining deep insights about what actions they wanted the agents to do, and what factors constituted their impression of the agents. Quantitative measures in isolation such as the score achieved, often only allow for discussions on whether they are significant or not.

Firstly, participants tended to be less forgiving of AI that obstructed their actions such as moving in front of them so that they wouldn't be able to move or place objects where they wanted. This would strongly offset any positive contributions the AI made throughout the levels as seen in Table 13. For example, an AI that performs optimal actions and adapts to the player well would receive negative sentiment if it ever obstructed the player, especially if it happened more than once. We found that this was a contributing factor to the larger fluctuations in cooperativity ratings the AA agent received, as when its partner's archetype was not clear, it would often collide with them due to sudden shifts in decision-making.

Table 13: Feedback about obstruction

| Data Source | Participant Quotes |
|---|---|
| Questionnaire | "The AI moved randomly around the workspace, blocking my character multiple times." |
| | "There were moments where it blocked my character from moving." |
| | "The AI moved randomly around the workspace, blocking my character multiple times." |
| | "Sometimes [the agent] often got in the way [and] blocked me when i tried to finish [and] help with the soup." |
| | "I often find that my partner blocks my way and usually doesn't move out of the way." |

Secondly, participants would not rate AI agents very highly in cooperativity even if the AI was productive in gaining a high score, as long as the AI did not make clear shifts in actions as a response to their actions. This is evident in quotes found in Table 14 where participants in our experiment mentioned changes in actions to be a factor in their perception of AI cooperativity. Furthermore, these shifts in actions would often usurp effective actions in priority for strong perceptions of cooperation, with participants preferring to see distinct adaptations rather than optimal cooperative actions.

Table 14: Feedback about adaptation

| Data Source | Participant Quotes |
|---|---|
| Questionnaire | "It doesn't feel like AI wants to work with me... I tried putting ingredients with them and they didn't even wait for me. It feels like the AI wants to do that role only so I adapted" |
| | "The bot did everything and didn't give me the chance to place food in the pot." |
| | "It did actions in sync with me at each stage of the game but we struggled at points with stuff needing to be done but nobody holding the right thing." |
| | "We swapped roles quite a bit but it worked well." |
| | "We had a good system (me on dishes, him on onions) but he changed roles at the end which slowed things down." |

### 7.1.2 Measuring Cooperative Motivation

Our experiment also made use of observational data which provided invaluable insights into patterns of participant behaviour, reasons why certain cooperating ratings were given, and why specific game scores were achieved. For example, using only quantitative measures alone, it would have been difficult to provide evidence for why the AA agent scores and cooperative ratings had larger variations compared to the other AI agents. However, with observational data, we were able to discover that due to certain strategies participants implemented, they would have very even archetypal mixtures which caused difficulties in the AA model, which would make drastic shifts in actions.

In addition, though it has not been explored thoroughly, we found that obstructions in pure human teams would not result in the same degree of negative sentiment compared to human AI teams. We believe this is the case as AI agents are not seen as part of the participant's 'group' and are instead seen as something to be subservient to them. This corroborates with past work on the nature of cooperation that describes it as the product of having agents with mutual goals, which in turn induces some form of joint action to achieve mutual goals and thus elicit some feeling of belonging [141, 171, 50, 14, 142]. AI agents are generally not perceived to have their own goals or intentions and thus do not elicit such a feeling. This hypothesis could be tested in other environments where the nature of a partner as a human or AI is hidden from participants.

### 7.1.3 Effective Cooperation Summary

For the reasons listed previously in the section, we believe that a combination of quantitative and qualitative measurements is needed to holistically assess the effectiveness of RL models in cooperating with human partners. Past work in cooperative AI in the RL space has largely avoided the use of qualitative measures despite frequent use outside of the area, in the broader field of cooperative AI and social studies.

These measures should be designed so that in combination together, they evaluate different factors of cooperation. In our experiment, we found the following trends as factors that contributed to the evaluation of AI agents by participants.

1. The result of cooperation: What impact does the AI agent have on the team score?
2. The perception of cooperation: Does the AI agent seem to act collaboratively? Are there attempts to adjust their actions to adapt to a partner?
3. The motivation of cooperation: Is there a desire for the AI agent to work with their partner? Does their partner want to work with them?

We do not believe these to be the complete list of factors that make up cooperation, but are ones that we found trends for, in our experiment.

## 7.2 On Mental Models of Humans

In line with expectations and other previous experiments [32], RL models that were trained with human data performed better in our experiments in both score and cooperative ratings compared to those that did not. It is, however, unclear whether this was due to being able to understand the intentions of their human partners and adapting to them respectively. We find that a more likely explanation for its better performance is due to being robust to more strategies that a partner may implement. This is because, during experiments, the adaptations they made for their partners were minimal and appeared more like a self-play model that was less focused on a singular role and was capable of fulfilling a supportive role. This robustness may be suitable to small environments where the possible approaches for cooperation are limited, such as the level of Overcooked we used for the experiment, but will likely struggle to be as effective in more complex environments. An example of this is the game 'Dota 2' where strategies are larger and involve the execution of smaller objectives over a long time horizon [129]. This implies that it is not entirely sufficient to simply use human data and expect that a meaningful understanding of different human strategies or personalities would be developed in RL models in the form of a state-action value function. Rather, it is necessary to construct a stronger understanding of player personalities to then be represented as part of the environment for effective cooperation to occur.

### 7.2.1 Reasons for Mental Models

Using a similar approach to Strouse et al. [162], combined with archetypal analysis, we found that a model without any human data could still achieve similar results in score and perceived cooperativity. This was likely the result of the aforementioned argument of robustness as the AA model was exposed to a wider range of strategies compared to the self-play model. Though our approach did not produce a model that performed significantly better than other benchmark models, a likely reason for this is that the environment we used to test these models did not put a strong emphasis on requiring quality teamwork between the agent and its partner to achieve a high score. It was perfectly possible to get a high score simply through efficient movement and actions, without any coordination with a partner. For example, when the player would deliver dishes then the AI agent would prepare the food, when the player prepared dishes then the AI agent would deliver dishes. Though there were some nuances such as the agent playing food on tables so that they were in more convenient locations for the player, the decisions were not overly difficult to replicate in a traditional state-machine-driven AI agent. As a result, there was less benefit to having a strong ability to adapt that our model had, and standard self-play models were sufficient to succeed in the environment.

Though the framework we used to develop the AA model did not make a significant improvement in the performance of the model compared to other models as measured by scoring, it did significantly improve how cooperative it was perceived by human partners due to its ability to make significant adaptations. Combined with the flexibility of our approach, it can easily serve as a wrapper for new and existing AI agents to improve adaptiveness. This is because our approach can effectively be condensed into computing AA on existing play data and using a partner's existing play data as a heuristic for informing decisions. This is not restricted to just RL models and can be used to improve traditional state-based models. An example of how this could be implemented is by computing the archetype mixture of a player and using that in conjunction with other environmental information to determine the appropriate action.

### 7.2.2 Archetypal Analysis for Developing Mental Models

Although it is likely the case that our approach does not result in the emergence of true cooperativity, we found that the use of clustering algorithms for the classification of player strategy does improve AI alignment and coordination with humans. The use of archetypal analysis to find innate patterns in a dataset of play data and effectively classify playstyles was a large component of this success. Though slower than comparable modern clustering techniques such as DBSCAN, AA aligns closer in function to work done in the field of psychology relating to human personality, including the Big Five and HEXACO models.

We worked around its large runtime complexity by running it offline and then passing its output archetypal profiles to an ensemble model as a heuristic for choosing actions of different cooperative models. This does limit our model to only infer the needs of a partner based on previous training data and means that if the complete range of playstyles were not captured during training, it may struggle to wholistically represent its partner's playstyle as a combination of existing archetypes. In addition, there is no guarantee that the model matches the partner's preferences or is aligned with them as their needs are imposed on them.

What our framework did succeed in was creating the perception of cooperation. This was through distinct shifts in actions in response to the strategy of the AI agent's partner. Evidence of this came from observational data, and feedback from questionnaires. Quantitatively, this did not have a significant impact on scores, but it did have an impact on the variation of scores, performing well for those with a strong archetypal alignment. We believe archetypal analysis was not a limiting factor and more work could be done in action selection using ensemble learning. We present opportunities for work in regard to these problems in chapter 8.4.

### 7.2.3 Mental Models Summary

For the reasons outlined in this section, we believe that producing better mental models of human partners in RL agents will improve their cooperativity.

Archetypal Analysis is the technique we used to achieve this, which believe, theoretically, is an effective approach as it creates a model that aligns with work in psychology research, which is also understandable for humans. The RL agent we developed using AA and ensemble learning proved to be capable of classifying human strategies and adapting to them, though suffered from drastic action shifts when partner archetypes were unclear. Despite this, the agent was developed with little additional cost compared to a typical self-play agent while significantly improving the adaptiveness to different partners, and how cooperative it was perceived.

## 7.3 Discussion Summary

In this chapter, we provided interpretations of the results of the experiment we ran, and summarised their implications. We first discussed the importance of combining quantitative and qualitative data for measuring the effective cooperativeness of RL agents. Next, we explored the value of developing mental models of human partners in AI agents for improving adaptiveness and perceived cooperativity. The use of Archetypal Analysis was justified for the construction of these models due to the similarities it shares with existing frameworks used to evaluate human personality.

# Chapter 8 Conclusion

We proposed a novel method that is simple to include in existing RL approaches and that has demonstrated promising results in making AI agents more cooperative with human partners. This was achieved by taking advantage of a clustering algorithm called archetypal analysis to better understand human partners and adapt to their actions. In doing so, we have established a promising direction for future research towards improving the cooperative capabilities of reinforcement learning models.

## 8.1 Thesis Overview

The thesis aimed to investigate whether the use of clustering techniques could be used to model human personality to improve the ability of AI models to cooperate with human partners. To guide our work, we proposed two research questions:

1. (RQ1) Is archetypal analysis suitable for use in a cooperative RL agent?
2. (RQ2) What methods are capable of measuring AI alignment with humans?

To address these questions, we developed a flexible framework to train a cooperative agent referred to as the 'AA agent' throughout the thesis, which was compared with benchmark models in a cooperative testbed 'Overcooked'. The framework works by finding archetypal playstyles in a dataset of playthrough data, training self-play models with each archetype to produce tailored policies for each, and then using ensemble learning to stochastically select the appropriate policies to cooperate with a novel human partner during run-time based on their archetypal mixture.

## 8.2 Contributions

In this section, we address the research questions proposed at the beginning of the thesis by summarising the conclusions that were reached as part of our experiments.

### 8.2.1 Research Question 1

(RQ1 - A) Archetypal analysis (AA) is effective for classifying player personalities and strategies and shows potential for use in the setting of cooperative RL agents. Based on our results, we found that our archetypal agent which was not trained on human data received higher levels of perceived cooperation by participants compared to self-play models and similar levels of perceived cooperation as RL models trained on human data. This was due to the agent demonstrating distinct shifts in actions depending on the user's personality and strategy, suggesting that formulating a better understanding of human partners enables higher levels of perceived cooperation.

(RQ1 - D) AA integration into the RL context is however limited by the algorithm's high computational complexity. To be integrated into an RL architecture, AA is too expensive to be run in real-time which means that the algorithm needs to be precomputed offline and will be unable to find new playstyles and strategies during run-time. As such, if not all archetypes are found from the initial dataset of playstyles that AA was run on, the algorithm may not be able to effectively model a human partner with a distinct, unseen personality. (RQ1 - B) Though this problem still exists, we were still able to successfully create an effective cooperative agent by using ensemble learning to switch between distinct cooperative policies deemed to be the best fit for a human partner's 'archetype'.

(RQ1 - C) These findings imply that clustering techniques such as AA can be used to model human profiles which can supplement the effectiveness of existing RL agents. Though we used these profiles to inform an ensemble learning architecture's selection of the appropriate cooperative policy, the human profiles developed through clustering could easily be used for any general RL agent. For example, instead of training two RL agents naively with one another, using clustering algorithms to improve their mental models of each other based on their actions could assist with their cooperativity.

### 8.2.2 Research Question 2

(RQ2 - A, B) Previous methods for comparing the cooperativity of RL models with human partners have typically come in the form of quantitative measures such as average rewards per episode [32] or the total score achieved by them [162]. However, these methods overlook important factors unique to cooperative contexts such as how cooperative they were perceived by and whether they made sufficient adaptations to their partners.

(RQ2 - C) To provide a more comprehensive understanding of the effectiveness of RL models in a cooperative setting, we integrated qualitative measures such as questionnaires and participant observations into our experiment. These were beneficial for assessing how participants felt about each RL model and extracted far more information about the cooperativity of each model compared to solely using quantitative metrics.

(RQ2 - D) The limitations of our approach in comparing the cooperativity of the agents are that reviews were generally overviews of the experience with few references to specific moments. This was because participants filled in the questionnaire after play and thus though they were fresh out of playing the game, it was difficult for them to refer to specific moments and discuss them in greater detail. We believe that by recording the playthroughs of the participants and stepping through different stages of their playthrough and discussing what they were thinking and expecting their AI partner to do, we would have been able to gather more informative results.

These findings imply that quantitative measures are insufficient for wholistically measuring how cooperative an AI agent is. Though metrics such as scores can provide indications for how well an RL model performs on its own, it does not have any implications on the cooperativeness of the model.

## 8.3 Limitations

This section outlines a few limitations of the work presented in this thesis.

### 8.3.1 Lack of access to benchmark data

Firstly, our ability to compare the experimental AA agent with benchmark models, such as the human-trained model, was limited due to the lack of access to the specific human data used to train the benchmark model. Although it was possible to develop our human-trained models using data of our own, we determined that our results would be more valuable and credible if they were directly comparable to models from established previous work.

### 8.3.2 Small number of participants

Secondly, our study was limited by the relatively small number of participants involved in the experiments. A larger sample size would have provided a more robust statistical foundation for our findings, allowing for a greater degree of confidence in our conclusions. Specifically, a larger participant pool would have enabled us to perform a more granular analysis of agent performance across different player skill levels. Throughout the experiments, we observed that more experienced players tended to rate both the human-trained and AA models more favorably, whereas less experienced players showed a preference for the self-play agent. However, due to the limited number of participants, particularly within each skill level, we were unable to draw definitive conclusions about these trends.

### 8.3.3 Limited environments

Additionally, our research was constrained by the limited number of environments in which we tested the performance of each agent. Ideally, testing across a wider variety of environments as well as more complex environments would provide deeper insights into the generalizability of the different agents. Such an approach could reveal additional strengths and weaknesses of the AA agent that may not be apparent in our work. However, the development and training of novel agents for multiple new environments require substantial computational resources and time that were not available to us during the project. This limitation restricted our ability to fully assess the versatility and adaptability of the AA agent across diverse scenarios.

## 8.4 Future Work

Our use of archetypal analysis as a heuristic for ensemble frameworks to better adapt to human partners is flexible, generalizable, and can easily be included in most existing model designs. However, throughout development, we found that there were a variety of directions that could be taken to improve the model.

### 8.4.1 Smoothing the transition between strategies

We found that a significant issue our model faced was the moments of indecision it faced when their partners did not align strongly with any archetypes. A potential approach to addressing this is by having the model stochastically choose a series of actions in the form of an action plan tailored towards a given archetype as opposed to a single action to perform.

### 8.4.2 Replacing self-play models with models trained with human proxies

For more simple levels, self-play agents appeared sufficient to generate diverse playstyles; however, for more complex environments and to create more unique strategies, the use of human data could greatly improve performance. As seen in Carroll et al. [32] there is no reason behaviour cloning could not be used to generate an improved dataset.

### 8.4.3 Increase or optimize the features observed

For the AA agent we trained for our experiment, we hand-picked features that we believed were the most important in identifying playstyles. This is, however, prone to error and is in no way optimal. Techniques could be implemented to better choose relevant features or a greater number of features could be tracked to generate more meaningful archetypes. As archetypal analysis also reduces the dimensions of the data, it should scale quite well with additional features.

### 8.4.4 More meaningful adaptations to non-archetypal partners

Currently, if partners do not fit into an archetype, the AA agent adapts by weighting the chances of different archetypal actions accordingly. However, this could done better. In human-human interactions, when cooperating with a partner, humans do not just randomly choose strategies that they believe their partner is in between. Therefore, AI agents should not do this either. Rather, new strategies often emerge, and so instead of weighting stochastic outcomes based on the partner's alignment with archetypes, the weights of the model could be scaled directly. This is an unexplored area and will likely require greater advancements in the field of explainable AI before anything significant can be achieved.

### 8.4.5 Faster clustering algorithms

AA and many other clustering techniques are quite slow in general and are not suitable for being computed in real time. This means that they may not be able to easily adapt to new playstyles that emerge during run-time. This currently makes gathering a diverse set of playstyles offline important to be effective at representing the player. Work in enhancing these algorithms to be able to run in real-time could overcome these limitations and make adaptation to emergent playstyles possible.

### 8.4.6 Shared mental models

The use of clustering demonstrated in this thesis only affects the AI understanding of the human partner while having minimal impact on the human understanding of the AI. Though this is a step towards cooperative AI, we believe that more work on developing a mutual understanding is needed. A simple display of the AI's current strategy could help inform their human partners more about them but we believe a more promising direction would be to pursue concrete ways of implementing shared mental models.

# Chapter 9 Appendix A

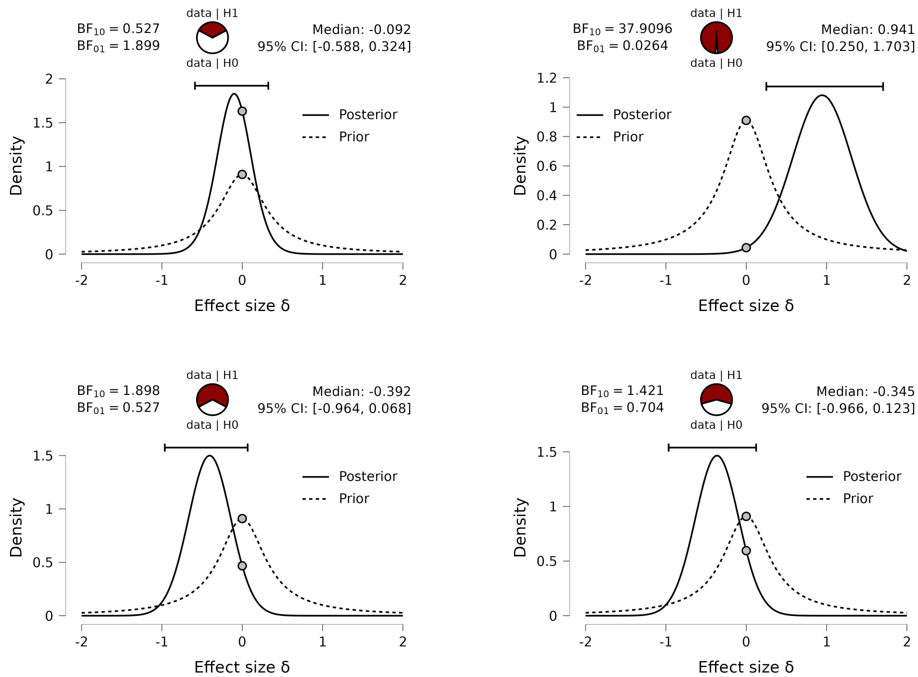


Figure 13: Inferential Plots of Bayesian Wilcoxon Signed-Rank Tests for Score: (Top-left) AA and Self-play, (Top-right) AA and Random, (Bot-left) AA and Human-trained, (Bot-right) Self-play and Human-Trained agents

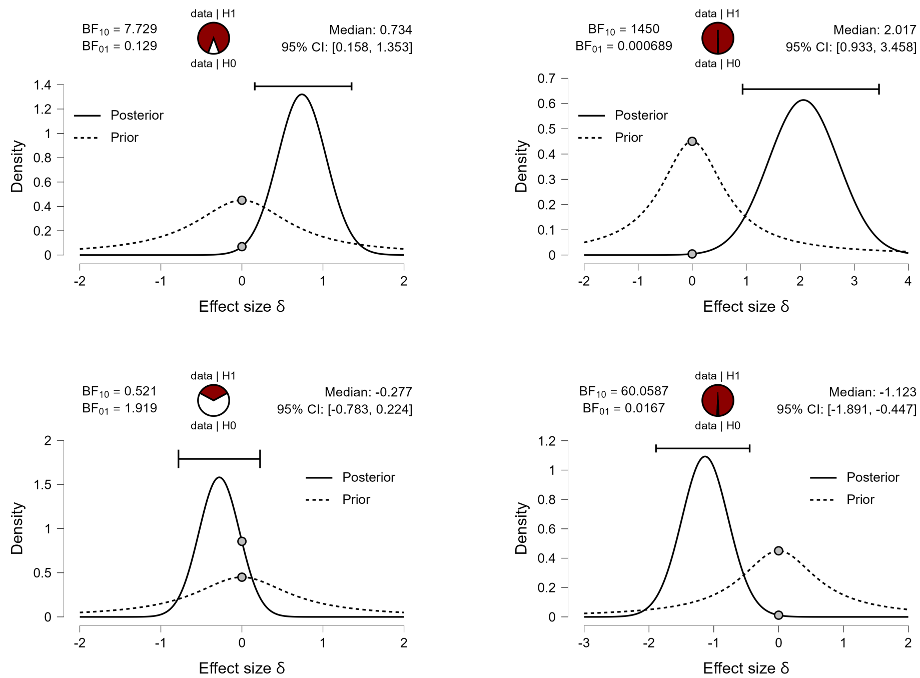


Figure 14: Inferential Plots of Bayesian Wilcoxon Signed-Rank Tests for Cooperative Rating: (Top-left) AA and Self-play, (Top-right) AA and Random, (Bot-left) AA and Human-trained, (Bot-right) Self-play and Human-Trained agents

# References


[1] Jalal Uddin Md Akbar et al. "A Comprehensive Review on Deep Learning Assisted Computer Vision Techniques for Smart Greenhouse Agriculture". In: *IEEE Access* 12 (2024). Conference Name: IEEE Access, pp. 4485–4522. ISSN: 2169-3536. DOI: `10.1109/ACCESS.2024.3349418`. URL: `https://ieeexplore.ieee.org/document/10379667` (visited on 06/02/2024).

[2] Victor Alvarez-Cortes et al. "Current Challenges and Applications for Adaptive User Interfaces". In: *Human-Computer Interaction* December (2009). DOI: 10.5772/7745.

[3] Carlos Alzate et al. "Identifying Customer Profiles in Power Load Time Series Using Spectral Clustering". en. In: *Artificial Neural Networks – ICANN 2009*. Ed. by Cesare Alippi et al. Berlin, Heidelberg: Springer, 2009, pp. 315–324. ISBN: 978-3-642-04277-5. DOI: 10.1007/978-3-642-04277-5_32.

[4] Daichi Amagata and Takahiro Hara. "Efficient Density-peaks Clustering Algorithms on Static and Dynamic Data in Euclidean Space". In: *ACM Transactions on Knowledge Discovery from Data* 18.1 (Aug. 2023), 2:1–2:27. ISSN: 1556-4681. DOI: 10.1145/3607873. URL: https://dl.acm.org/doi/10.1145/3607873 (visited on 04/16/2024).

[5] Saleema Amershi et al. "Guidelines for Human-AI Interaction". en. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow Scotland Uk: ACM, May 2019, pp. 1–13. ISBN: 978-1-4503-5970-2. DOI: 10.1145/3290605.3300233. URL: https://dl.acm.org/doi/10.1145/3290605.3300233 (visited on 01/26/2022).

[6] Robert W. Andrews et al. "The role of shared mental models in human-AI teams: a theoretical review". In: *Theoretical Issues in Ergonomics Science* 24.2 (Mar. 2023), pp. 129–175. ISSN: 1463-922X. DOI: 10.1080/1463922X.2022.2061080. URL: https://doi.org/10.1080/1463922X.2022.2061080 (visited on 04/20/2024).

[7] Saurabh Arora and Prashant Doshi. "A survey of inverse reinforcement learning: Challenges, methods and progress". In: *Artificial Intelligence* 297 (Aug. 2021). Publisher: Elsevier B.V. _eprint: 1806.06877, p. 103500. ISSN: 00043702. DOI: 10.1016/j.artint.2021.103500. URL: https://doi.org/10.1016/j.artint.2021.103500%20https://linkinghub.elsevier.com/retrieve/pii/S0004370221000515.

[8] Christian Arzate Cruz and Takeo Igarashi. "MarioMix: Creating aligned playstyles for bots with interactive reinforcement learning". In: *CHI PLAY 2020 - Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play* (2020). ISBN: 9781450375870, pp. 134–139. DOI: 10.1145/3383668.3419938.

[9] Zahra Ashktorab et al. "Human-AI Collaboration in a Cooperative Game Setting". In: *Proceedings of the ACM on Human-Computer Interaction* 4.CSCW2 (2020), pp. 1–20. ISSN: 2573-0142. DOI: 10.1145/3415167.

[10] Zahra Ashktorab et al. "Human-AI Collaboration in a Cooperative Game Setting". In: *Proceedings of the ACM on Human-Computer Interaction* 4.CSCW2 (2020), pp. 1–20. ISSN: 2573-0142. DOI: 10.1145/3415167.

[11] Michael C. Ashton and Kibeom Lee. "Empirical, Theoretical, and Practical Advantages of the HEXACO Model of Personality Structure". en. In: *Personality and Social Psychology Review* 11.2 (May 2007). Publisher: SAGE Publications Inc, pp. 150–166. ISSN: 1088-8683. DOI: 10.1177/1088868306294907. URL: https://doi.org/10.1177/1088868306294907 (visited on 04/21/2024).

[12] Michael C. Ashton and Kibeom Lee. "The HEXACO Model of Personality Structure and the Importance of the H Factor". en. In: *Social and Personality Psychology Compass* 2.5 (2008), pp. 1952–1962. ISSN: 1751-9004. DOI: 10.1111/j.1751-9004.2008.00134.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1751-9004.2008.00134.x (visited on 04/21/2024).

[13] Michael C. Ashton and Kibeom Lee. "The HEXACO–60: A Short Measure of the Major Dimensions of Personality". In: *Journal of Personality Assessment* 91.4 (July 2009), pp. 340–345. ISSN: 0022-3891. DOI: 10.1080/00223890902935878. URL: https://www.tandfonline.com/doi/abs/10.1080/00223890902935878 (visited on 04/21/2024).

[14] Robert Axelrod. “On Six Advances in Cooperation Theory”. de. In: *Analyse & Kritik* 22.1 (May 2000). Publisher: De Gruyter Oldenbourg, pp. 130–151. ISSN: 2365-9858. DOI: 10.1515/auk-2000-0107. URL: https://www.degruyter.com/document/doi/10.1515/auk-2000-0107/html (visited on 12/05/2021).

[15] R. Michael Bagby and Thomas A. Widiger. “Five Factor Model personality disorder scales: An introduction to a special section on assessment of maladaptive variants of the five factor model”. eng. In: *Psychological Assessment* 30.1 (Jan. 2018), pp. 1–9. ISSN: 1939-134X. DOI: 10.1037/pas0000523.

[16] Nathan R. Bailey et al. “A Brain-Based Adaptive Automation System and Situation Awareness: The Role of Complacency Potential”. en. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 47.9 (Oct. 2003). Publisher: SAGE Publications Inc, pp. 1048–1052. ISSN: 1071-1813. DOI: 10.1177/154193120304700901. URL: https://doi.org/10.1177/154193120304700901 (visited on 04/21/2024).

[17] Bowen Baker et al. “Emergent Tool Use From Multi-Agent Autocurricula”. In: *arXiv:1909.07528 [cs, stat]* (Feb. 2020). arXiv: 1909.07528. URL: http://arxiv.org/abs/1909.07528 (visited on 11/07/2021).

[18] Gagan Bansal et al. “Beyond Accuracy: The Role of Mental Models in Human-AI Team Performance”. en. In: *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 7 (Oct. 2019), pp. 2–11. URL: https://ojs.aaai.org/index.php/HCOMP/article/view/5285 (visited on 01/26/2022).

[19] Gagan Bansal et al. “Updates in Human-AI Teams: Understanding and Addressing the Performance/Compatibility Tradeoff”. en. In: *Proceedings of the AAAI Conference on Artificial Intelligence* 33.01 (July 2019). Number: 01, pp. 2429–2437. ISSN: 2374-3468. DOI: 10.1609/aaai.v33i01.33012429. URL: https://ojs.aaai.org/index.php/AAAI/article/view/4087 (visited on 01/26/2022).

[20] Nolan Bard et al. "The Hanabi challenge: A new frontier for AI research". In: *Artificial Intelligence* 280 (2020). _eprint: 1902.00506. ISSN: 00043702. DOI: 10.1016/j.artint.2019.103216.

[21] Joseph Bates. "The Role of Emotion in Believable Agents". In: *Communications of the ACM* 37.7 (1994), pp. 122–125. ISSN: 15577317. DOI: 10.1145/176789.176803.

[22] Christian Bauckhage, Anders Drachen, and Rafet Sifa. "Clustering Game Behavior Data". en. In: *IEEE Transactions on Computational Intelligence and AI in Games* 7.3 (Sept. 2015), pp. 266–278. ISSN: 1943-068X, 1943-0698. DOI: 10.1109/TCIAIG.2014.2376982. URL: https://ieeexplore.ieee.org/document/6975073/ (visited on 10/10/2021).

[23] Christian Bauckhage and Christian Thurau. "Making archetypal analysis practical". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 5748 LNCS (2009). ISBN: 3642037976, pp. 272–281. ISSN: 03029743. DOI: 10.1007/978-3-642-03798-6_28.

[24] Neel Jitesh Bhate et al. "Zero-shot Learning with Minimum Instruction to Extract Social Determinants and Family History from Clinical Notes using GPT Model". In: *2023 IEEE International Conference on Big Data (BigData)*. Dec. 2023, pp. 1476–1480. DOI: 10.1109/BigData59044.2023.10386811. URL: https://ieeexplore.ieee.org/document/10386811 (visited on 06/02/2024).

[25] Filippo Maria Bianchi, Daniele Grattarola, and Cesare Alippi. "Spectral clustering with graph neural networks for graph pooling". In: *Proceedings of the 37th International Conference on Machine Learning*. Vol. 119. ICML'20. JMLR.org, July 2020, pp. 874–883. (Visited on 04/18/2024).

[26] Temi Bidjerano and David Yun Dai. "The relationship between the big-five model of personality and self-regulated learning strategies". In: *Learning and Individual Differences* 17.1 (Jan. 2007), pp. 69–81. ISSN: 1041-6080. DOI: 10.1016/j.lindif.2007.02.001. URL: https://www.

sciencedirect.com/science/article/pii/S104160800700026X (visited on 04/21/2024).

[27] Athman Bouguettaya et al. "Efficient agglomerative hierarchical clustering". In: *Expert Systems with Applications* 42.5 (Apr. 2015), pp. 2785–2797. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2014.09.054. URL: https://www.sciencedirect.com/science/article/pii/S0957417414006150 (visited on 04/22/2024).

[28] Christos Boutsidis, Alex Gittens, and Prabhanjan Kambadur. "Spectral clustering via the power method - provably". In: *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*. ICML'15. Lille, France: JMLR.org, July 2015, pp. 40–48. (Visited on 04/18/2024).

[29] T Brunélis et al. "Distant mission management and dynamic allocation of functions". In: *Human Operating Unmanned Systems (HUMOUS)* (Sept. 2008).

[30] Kris Cao et al. "Emergent communication through negotiation". In: *arXiv* (2018). _eprint: 1804.03980, pp. 1–15. ISSN: 23318422.

[31] Micah Carroll. "Overview of current AI Alignment Approaches". In: December (2018).

[32] Micah Carroll et al. "On the utility of learning about humans for human-AI coordination". In: *arXiv* NeurIPS (2019). _eprint: 1910.05789. ISSN: 23318422.

[33] Tristan Cazenave. "Residual Networks for Computer Go". In: *IEEE Transactions on Games* 10.1 (Mar. 2018). Conference Name: IEEE Transactions on Games, pp. 107–110. ISSN: 2475-1510. DOI: 10.1109/TCIAIG.2017.2681042.

[34] Urszula Chajewska, Daphne Koller, and Dirk Ormoneit. "Learning an Agent's Utility Function by Observing Behaviour". In: (2005).

[35] B. H. P. Chan, D. A. Mitchell, and L. E. Cram. “Archetypal analysis of galaxy spectra”. en. In: *Monthly Notices of the Royal Astronomical Society* 338.3 (Jan. 2003), pp. 790–795. ISSN: 00358711, 13652966. DOI: 10.1046/j.1365-8711.2003.06099.x. URL: https://academic.oup.com/mnras/article-lookup/doi/10.1046/j.1365-8711.2003.06099.x (visited on 10/13/2021).

[36] Prithvijit Chattopadhyay et al. “Evaluating visual conversational agents via cooperative human-AI games”. In: *arXiv* Hcomp (2017). _eprint: 1708.05122, pp. 2–10. ISSN: 23318422.

[37] Bi Yu Chen et al. “Efficient and scalable DBSCAN framework for clustering continuous trajectories in road networks”. In: *International Journal of Geographical Information Science* 37.8 (Aug. 2023). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/13658816.2023.2217443, pp. 1693–1727. ISSN: 1365-8816. DOI: 10.1080/13658816.2023.2217443. URL: https://doi.org/10.1080/13658816.2023.2217443 (visited on 04/19/2024).

[38] Yuansi Chen, Julien Mairal, and Zaid Harchaoui. “Fast and robust archetypal analysis for representation learning”. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2014). ISBN: 9781479951178 _eprint: 1405.6472, pp. 1478–1485. ISSN: 10636919. DOI: 10.1109/CVPR.2014.192.

[39] Elena Denny Cherpanath et al. “Food Image Recognition and Calorie Prediction Using Faster R-CNN and Mask R-CNN”. In: *2023 9th International Conference on Smart Computing and Communications (ICSCC)*. Aug. 2023, pp. 83–89. DOI: 10.1109/ICSCC59169.2023.10335053. URL: https://ieeexplore.ieee.org/document/10335053 (visited on 06/02/2024).

[40] Yann Chevaleyre et al. “A Short Introduction to Computational Social Choice”. en. In: *SOFSEM 2007: Theory and Practice of Computer Science*. Ed. by Jan van Leeuwen et al. Lecture Notes in Computer Science.

Berlin, Heidelberg: Springer, 2007, pp. 51–69. ISBN: 978-3-540-69507-3. DOI: 10.1007/978-3-540-69507-3_4.

[41] Eugenio Chisari et al. "Correct Me If I am Wrong: Interactive Learning for Robotic Manipulation". en. In: *IEEE Robotics and Automation Letters* 7.2 (Apr. 2022), pp. 3695–3702. ISSN: 2377-3766, 2377-3774. DOI: 10.1109/LRA.2022.3145516. URL: https://ieeexplore.ieee.org/document/9691826/ (visited on 04/21/2024).

[42] Robin Cohen et al. "Trusted Al and the contribution of trust modeling in multiagent systems". In: *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS* 3 (2019). ISBN: 9781510892002, pp. 1644–1648. ISSN: 15582914.

[43] Vincent Cohen-addad et al. "Hierarchical Clustering: Objective Functions and Algorithms". In: *Journal of the ACM* 66.4 (June 2019), 26:1–26:42. ISSN: 0004-5411. DOI: 10.1145/3321386. URL: https://dl.acm.org/doi/10.1145/3321386 (visited on 04/20/2024).

[44] Vincent Conitzer and Tuomas Sandholm. "Vote elicitation: Complexity and strategy-proofness". In: *Proceedings of the National Conference on Artificial Intelligence* (2002), pp. 392–397. DOI: 10.1184/r1/6612824.

[45] Paul T. Costa Jr. and Thomas A. Widiger, eds. *Personality disorders and the five-factor model of personality, 2nd ed.* Personality disorders and the five-factor model of personality, 2nd ed. Pages: xii, 493. Washington, DC, US: American Psychological Association, 2002. ISBN: 978-1-55798-826-3. DOI: 10.1037/10423-000.

[46] Christian Arzate Cruz and Jorge Adolfo Ramirez Uresti. *HRLB^2: A reinforcement learning based framework for believable bots.* ISSN: 20763417 Issue: 12 Publication Title: Applied Sciences (Switzerland) Volume: 8. 2018. DOI: 10.3390/app8122453.

[47] Adele Cutler and Leo Breiman. "Archetypal analysis". In: *Technometrics* 36.4 (1994), pp. 338–347. ISSN: 15372723. DOI: 10.1080/00401706.1994.10485840.

[48] Abolfazl Dadgarnia and Mohammad Taghi Sadeghi. "A novel method of deinterleaving radar pulse sequences based on a modified DBSCAN algorithm". In: *China Communications* 20.2 (Feb. 2023). Conference Name: China Communications, pp. 198–215. ISSN: 1673-5447. DOI: 10.23919/JCC.2023.02.012. URL: https://ieeexplore.ieee.org/document/10061671 (visited on 04/19/2024).

[49] Allan Dafoe et al. "Cooperative AI: machines must learn to find common ground". In: *Nature 2021 593:7857* 593.7857 (May 2021). Publisher: Nature Publishing Group, pp. 33–36. DOI: 10.1038/d41586-021-01170-0. URL: https://www.nature.com/articles/d41586-021-01170-0.

[50] Allan Dafoe et al. "Open problems in cooperative AI". In: *arXiv* August 2019 (2020). _eprint: 2012.08630. ISSN: 23318422.

[51] Anil Damle and Yuekai Sun. "A Geometric Approach to Archetypal Analysis and Nonnegative Matrix Factorization". In: *Technometrics* 59.3 (2017). _eprint: 1405.4275, pp. 361–370. ISSN: 15372723. DOI: 10.1080/00401706.2016.1247017.

[52] William H. E. Day and Herbert Edelsbrunner. "Efficient algorithms for agglomerative hierarchical clustering methods". en. In: *Journal of Classification* 1.1 (Dec. 1984), pp. 7–24. ISSN: 1432-1343. DOI: 10.1007/BF01890115. URL: https://doi.org/10.1007/BF01890115 (visited on 04/22/2024).

[53] Hinton Dayan. "Feudal Reinforcement Learning". en. In: (), p. 8.

[54] Ezequiel Debada, Laleh Makarem, and Denis Gillet. "Autonomous coordination of heterogeneous vehicles at roundabouts". In: *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. ISSN: 2153-0017. Nov. 2016, pp. 1489–1495. DOI: 10.1109/ITSC.2016.

7795754. URL: https://ieeexplore.ieee.org/document/7795754 (visited on 04/20/2024).

[55] Soledad Delgado et al. "Analysis of Students' Behavior Through User Clustering in Online Learning Settings, Based on Self Organizing Maps Neural Networks". In: *IEEE Access* 9 (2021). Conference Name: IEEE Access, pp. 132592–132608. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3115024. URL: https://ieeexplore.ieee.org/document/9546766 (visited on 04/21/2024).

[56] Song Deng et al. "User Behavior Analysis Based on Stacked Autoencoder and Clustering in Complex Power Grid Environment". In: *IEEE Transactions on Intelligent Transportation Systems* 23.12 (Dec. 2022). Conference Name: IEEE Transactions on Intelligent Transportation Systems, pp. 25521–25535. ISSN: 1558-0016. DOI: 10.1109/TITS.2021.3076607. URL: https://ieeexplore.ieee.org/document/9430771 (visited on 04/20/2024).

[57] Hu Ding, Fan Yang, and Mingyue Wang. "On metric DBSCAN with low doubling dimension". In: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*. IJCAI'20. Yokohama, Yokohama, Japan, Jan. 2021, pp. 3080–3086. ISBN: 978-0-9992411-6-5. (Visited on 04/18/2024).

[58] Zihan Ding. "Imitation Learning: A Survey of Learning Methods". In: *Deep Reinforcement Learning: Fundamentals, Research and Applications* 50.2 (2020). ISBN: 9789811540950, pp. 273–306. DOI: 10.1007/978-981-15-4095-0_8.

[59] Julie Doron et al. "Coping profiles, perceived stress and health-related behaviors: a cluster analysis approach". eng. In: *Health Promotion International* 30.1 (Mar. 2015), pp. 88–100. ISSN: 1460-2245. DOI: 10.1093/heapro/dau090.

[60] Anders Drachen, Alessandro Canossa, and Georgios N. Yannakakis. "Player modeling using self-organization in Tomb Raider: Underworld". en. In: *2009 IEEE Symposium on Computational Intelligence and Games*. Milano, Italy: IEEE, Sept. 2009, pp. 1–8. ISBN: 978-1-4244-4814-2. DOI: 10.1109/CIG.2009.5286500. URL: http://ieeexplore.ieee.org/document/5286500/ (visited on 01/26/2022).

[61] Anders Drachen et al. *A Comparison of Methods for Player Clustering via Behavioral Telemetry*. arXiv:1407.3950 [cs]. July 2014. DOI: 10.48550/arXiv.1407.3950. URL: http://arxiv.org/abs/1407.3950 (visited on 04/20/2024).

[62] Anders Drachen et al. "Guns, swords and data: Clustering of player behavior in computer games in the wild". en. In: *2012 IEEE Conference on Computational Intelligence and Games (CIG)*. Granada, Spain: IEEE, Sept. 2012, pp. 163–170. ISBN: 978-1-4673-1193-9 978-1-4673-1194-6. DOI: 10.1109/CIG.2012.6374152. URL: https://ieeexplore.ieee.org/document/6374152/ (visited on 04/20/2024).

[63] John J. Dudley and Per Ola Kristensson. "A review of user interface design for interactive machine learning". In: *ACM Transactions on Interactive Intelligent Systems* 8.2 (2018). ISSN: 21606463. DOI: 10.1145/3185517.

[64] John J. Dudley et al. "A review of user interface design for interactive machine learning". In: *DIS 2020 - Proceedings of the 2020 ACM Designing Interactive Systems Conference* 8.2 (2020). ISBN: 9781450369749, pp. 1195–1209. ISSN: 21606463. DOI: 10.1145/3185517.

[65] Mustafa Suphi Erden and Tetsuo Tomiyama. "Human-Intent Detection and Physically Interactive Control of a Robot Without Force Sensors". In: *IEEE Transactions on Robotics* 26.2 (Apr. 2010). Conference Name: IEEE Transactions on Robotics, pp. 370–382. ISSN: 1941-0468. DOI: 10.1109/TRO.2010.2040202.

[66] Manuel J. A. Eugster. “Performance Profiles based on Archetypal Athletes”. en. In: *International Journal of Performance Analysis in Sport* 12.1 (Apr. 2012), pp. 166–187. ISSN: 2474-8668, 1474-8185. DOI: 10.1080/24748668.2012.11868592. URL: https://www.tandfonline.com/doi/full/10.1080/24748668.2012.11868592 (visited on 10/11/2021).

[67] Shi Feng and Jordan Boyd-Graber. “What can AI do for me?” In: (2019). ISBN: 9781450362726, pp. 229–239. DOI: 10.1145/3301275.3302265.

[68] Shi Feng et al. “Learning an Agent’s Utility Function by Observing Behaviour”. In: *Proceedings of the ACM on Human-Computer Interaction* 39.1 (2020). ISBN: 9781450362726, pp. 1–12. ISSN: 2573-0142. DOI: 10.1145/3415167. URL: https://www.researchgate.net/profile/Karel_Bosch/publication/325718292_Human-AI_Cooperation_to_Benefit_Military_Decision_Making/links/5b1fc0b2a6fdcc69745cd169/Human-AI-Cooperation-to-Benefit-Military-Decision-Making.pdf.

[69] Daniel Fernández, Irene Epifanio, and Louise McMillan. “Archetypal Analysis for Ordinal Data”. In: *Information Sciences* 579 (Aug. 2021). DOI: 10.1016/j.ins.2021.07.095.

[70] Chelsea Finn, Pieter Abbeel, and Sergey Levine. “Model-agnostic meta-learning for fast adaptation of deep networks”. In: *arXiv* (2017). ISSN: 23318422.

[71] Carlos Florensa, Yan Duan, and Pieter Abbeel. “Stochastic Neural Networks for Hierarchical Reinforcement Learning”. In: *arXiv:1704.03012 [cs]* (Apr. 2017). arXiv: 1704.03012 version: 1. URL: http://arxiv.org/abs/1704.03012 (visited on 02/26/2022).

[72] Vincent Francois-Lavet et al. “An Introduction to Deep Reinforcement Learning”. In: *Foundations and Trends® in Machine Learning* 11.3-4 (2018). arXiv: 1811.12560, pp. 219–354. ISSN: 1935-8237, 1935-8245. DOI: 10.1561/2200000071. URL: http://arxiv.org/abs/1811.12560 (visited on 02/26/2022).

[73] Benjamin Fung, ke Wang, and Martin Ester. "Hierarchical Document Clustering". In: vol. 1. Jan. 2006. DOI: 10.4018/9781591405573.ch105.

[74] Katy Ilonka Gero et al. "Mental Models of AI Agents in a Cooperative Game Setting". In: *Conference on Human Factors in Computing Systems - Proceedings* (2020). ISBN: 9781450367080, pp. 1–12. DOI: 10.1145/3313831.3376316.

[75] Felix Gervits et al. "How Should Agents Ask Questions For Situated Learning? An Annotated Dialogue Corpus". en. In: *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Singapore and Online: Association for Computational Linguistics, 2021, pp. 353–359. DOI: 10.18653/v1/2021.sigdial-1.37. URL: https://aclanthology.org/2021.sigdial-1.37 (visited on 04/21/2024).

[76] A. Guénoche, P. Hansen, and B. Jaumard. "Efficient algorithms for divisive hierarchical clustering with the diameter criterion". en. In: *Journal of Classification* 8.1 (Jan. 1991), pp. 5–30. ISSN: 1432-1343. DOI: 10.1007/BF02616245. URL: https://doi.org/10.1007/BF02616245 (visited on 04/22/2024).

[77] Wenjie Guo et al. "Density Peak Clustering with connectivity estimation". In: *Knowledge-Based Systems* 243 (May 2022), p. 108501. ISSN: 0950-7051. DOI: 10.1016/j.knosys.2022.108501. URL: https://www.sciencedirect.com/science/article/pii/S0950705122002155 (visited on 04/16/2024).

[78] Jayesh K. Gupta, Maxim Egorov, and Mykel Kochenderfer. "Cooperative Multi-agent Control Using Deep". In: *Autonomous Agents and Multiagent Systems* (2017). ISBN: 9783319716824, pp. 66–83. DOI: 10.1007/978-3-319-71682-4.

[79] Dylan Hadfield-Menell et al. "Cooperative inverse reinforcement learning". In: *Advances in Neural Information Processing Systems* Nips (2016). _eprint: 1606.03137, pp. 3916–3924. ISSN: 10495258.

[80] Nader Hanna and Deborah Richards. "The Impact of Multimodal Communication on a Shared Mental Model, Trust, and Commitment in Human–Intelligent Virtual Agent Teams". en. In: *Multimodal Technologies and Interaction* 2.3 (Aug. 2018), p. 48. ISSN: 2414-4088. DOI: 10.3390/mti2030048. URL: http://www.mdpi.com/2414-4088/2/3/48 (visited on 04/21/2024).

[81] Peter Henderson et al. "Deep Reinforcement Learning that Matters". In: *arXiv:1709.06560 [cs, stat]* (Jan. 2019). arXiv: 1709.06560. URL: http://arxiv.org/abs/1709.06560 (visited on 02/26/2022).

[82] Philip Hingston. "A turing test for computer game bots". In: *IEEE Transactions on Computational Intelligence and AI in Games* 1.3 (2009), pp. 169–186. ISSN: 1943068X. DOI: 10.1109/TCIAIG.2009.2032534.

[83] Daniel Ho et al. "Population Based Augmentation: Efficient Learning of Augmentation Policy Schedules". In: *arXiv:1905.05393 [cs, stat]* (May 2019). arXiv: 1905.05393 version: 1. URL: http://arxiv.org/abs/1905.05393 (visited on 02/26/2022).

[84] Jonathan Ho and Stefano Ermon. "Generative adversarial imitation learning". In: *Advances in Neural Information Processing Systems* (2016). _eprint: 1606.03476, pp. 4572–4580. ISSN: 10495258.

[85] Steven Holtzen et al. "Inferring human intent from video by sampling hierarchical plans". In: *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. ISSN: 2153-0866. Oct. 2016, pp. 1489–1496. DOI: 10.1109/IROS.2016.7759242.

[86] Borja Ibarz et al. "Reward learning from human preferences and demonstrations in Atari". In: *arXiv* 2017 (2018), pp. 1–13. ISSN: 23318422.

[87] Walaa N. Ismail et al. "CNN-Based Health Model for Regular Health Factors Analysis in Internet-of-Medical Things Environment". In: *IEEE Access* 8 (2020). Conference Name: IEEE Access, pp. 52541–52549. ISSN:

2169-3536. DOI: 10 . 1109 / ACCESS . 2020 . 2980938. URL: https : / / ieeexplore.ieee.org/document/9036869 (visited on 06/02/2024).

[88] Max Jaderberg et al. "Human-level performance in 3D multiplayer games with population- based reinforcement learning". In: 865.May (2019), pp. 859–865.

[89] Max Jaderberg et al. "Population based training of neural networks". In: *arXiv* (2017). _eprint: 1711.09846. ISSN: 23318422.

[90] Natasha Jaques et al. "Social influence as intrinsic motivation for multi-agent deep reinforcement learning". In: *36th International Conference on Machine Learning, ICML 2019* 2019-June (2019). ISBN: 9781510886988 _eprint: 1810.08647, pp. 5372–5381.

[91] Oumayma Jouini, Kaouthar Sethom, and Ridha Bouallegue. "Wheat leaf disease detection using CNN in Smart Agriculture". In: *2023 International Wireless Communications and Mobile Computing (IWCMC)*. ISSN: 2376-6506. June 2023, pp. 1660–1665. DOI: 10.1109/IWCMC58020.2023.10183348. URL: https://ieeexplore.ieee.org/document/10183348 (visited on 06/02/2024).

[92] Timothy A. Judge, Daniel Heller, and Michael K. Mount. "Five-factor model of personality and job satisfaction: A meta-analysis". In: *Journal of Applied Psychology* 87.3 (2002). Place: US Publisher: American Psychological Association, pp. 530–541. ISSN: 1939-1854. DOI: 10.1037/0021-9010.87.3.530.

[93] David B. Kaber and Mica R. Endsley. "The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task". In: *Theoretical Issues in Ergonomics Science* 5.2 (2004). Place: United Kingdom Publisher: Taylor & Francis, pp. 113–153. ISSN: 1464-536X. DOI: 10.1080/1463922021000054335.

[94] David B. Kaber et al. "Adaptive Automation of Human-Machine System Information-Processing Functions". en. In: *Human Factors* 47.4 (Dec.

2005). Publisher: SAGE Publications Inc, pp. 730–741. ISSN: 0018-7208. DOI: 10.1518/001872005775570989. URL: https://doi.org/10.1518/001872005775570989 (visited on 04/21/2024).

[95] Yoko Kadowaki et al. "Development of Soft Power-Assist Glove and Control Based on Human Intent". In: *Journal of Robotics and Mechatronics* 23.2 (Apr. 2011). Publisher: Fuji Technology Press Ltd., pp. 281–291. DOI: 10.20965/jrm.2011.p0281. URL: https://www.fujipress.jp/jrm/rb/robot002300020281/ (visited on 02/22/2022).

[96] Tushar Kansal et al. "Customer Segmentation using K-means Clustering". In: *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)*. Dec. 2018, pp. 135–139. DOI: 10.1109/CTEMS.2018.8769171. URL: https://ieeexplore.ieee.org/abstract/document/8769171 (visited on 05/11/2024).

[97] Sebastian Mathias Keller et al. "Deep Archetypal Analysis". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11824 LNCS (2019). ISBN: 9783030336752 _eprint: 1901.10799, pp. 171–185. ISSN: 16113349. DOI: 10.1007/978-3-030-33676-9_12.

[98] Shauharda Khadka and Kagan Tumer. "Evolution-Guided Policy Gradient in Reinforcement Learning". In: *arXiv:1805.07917 [cs, stat]* (Oct. 2018). arXiv: 1805.07917. URL: http://arxiv.org/abs/1805.07917 (visited on 02/26/2022).

[99] Teuvo Kohonen. "Self-Organizing Maps". In: ed. by Thomas S. Huang et al. Vol. 30. Springer Series in Information Sciences. Berlin, Heidelberg: Springer Berlin Heidelberg, 1995. ISBN: 978-3-642-97612-4 978-3-642-97610-0. DOI: 10.1007/978-3-642-97610-0. URL: http://link.springer.com/10.1007/978-3-642-97610-0 (visited on 05/02/2024).

[100] Talia Lavie and Joachim Meyer. "Benefits and costs of adaptive user interfaces". In: *International Journal of Human-Computer Studies*. Measuring the Impact of Personalization and Recommendation on User Be-

haviour 68.8 (Aug. 2010), pp. 508–524. ISSN: 1071-5819. DOI: 10.1016/j.ijhcs.2010.01.004. URL: https://www.sciencedirect.com/science/article/pii/S1071581910000145 (visited on 04/21/2024).

[101] Jang-Hwan Lee and Young-Gul Kim. "A stage model of organizational knowledge management: A latent content analysis". In: *Expert Systems with Applications* 20 (May 2001), pp. 299–311. DOI: 10.1016/S0957-4174(01)00015-X.

[102] Jan Leike et al. "Scalable agent alignment via reward modeling: A research direction". In: *arXiv* (2018). _eprint: 1811.07871. ISSN: 23318422.

[103] Adam Lerer and Alexander Peysakhovich. "Learning existing social conventions via observationally augmented self-play". In: *AIES 2019 - Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (2019). ISBN: 9781450363242 _eprint: 1806.10071, pp. 107–114. DOI: 10.1145/3306618.3314268.

[104] Adam Lerer et al. "MarioMix: Creating aligned playstyles for bots with interactive reinforcement learning". In: *arXiv* 9.NeurIPS (2020). ISBN: 9781450375184 Publisher: Springer US _eprint: 1703.06207, pp. 134–139. ISSN: 23318422. DOI: 10.1609/aaai.v34i05.6208. URL: http://dx.doi.org/10.1038/s41467-017-02597-8.

[105] Shan Li et al. "ARCHETYPAL ANALYSIS: A NEW WAY TO SEGMENT MARKETS BASED ON EXTREME INDIVIDUALS". en. In: (2003), p. 6.

[106] Jin Lian. "Implementation of computer network user behavior forensic analysis system based on speech data system log". en. In: *International Journal of Speech Technology* 23.3 (Sept. 2020), pp. 559–567. ISSN: 1572-8110. DOI: 10.1007/s10772-020-09747-2. URL: https://doi.org/10.1007/s10772-020-09747-2 (visited on 04/21/2024).

[107] Hongfu Liu, Zhiqiang Tao, and Yun Fu. "Partition Level Constrained Clustering". In: *IEEE Transactions on Pattern Analysis and Machine*

*Intelligence* 40.10 (Oct. 2018). Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 2469–2483. ISSN: 1939-3539. DOI: 10.1109/TPAMI.2017.2763945. URL: https://ieeexplore.ieee.org/document/8070337 (visited on 04/20/2024).

[108] Daniel Livingstone. "Turing's test and believable AI in games". In: *Computers in Entertainment* 4.1 (2006), pp. 31–43. ISSN: 15443574. DOI: 10.1145/1111293.1111303.

[109] Dylan P. Losey et al. "A Review of Intent Detection, Arbitration, and Communication Aspects of Shared Control for Physical Human–Robot Interaction". In: *Applied Mechanics Reviews* 70.1 (Feb. 2018). ISSN: 0003-6900. DOI: 10.1115/1.4039145. URL: https://doi.org/10.1115/1.4039145 (visited on 02/22/2022).

[110] J. MacQueen. "Some methods for classification and analysis of multivariate observations". In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics.* Vol. 5.1. University of California Press, Jan. 1967, pp. 281–298. URL: https://projecteuclid.org/ebooks/berkeley-symposium-on-mathematical-statistics-and-probability/Proceedings-of-the-Fifth-Berkeley-Symposium-on-Mathematical-Statistics-and/chapter/Some-methods-for-classification-and-analysis-of-multivariate-observations/bsmsp/1200512992 (visited on 04/20/2024).

[111] Sebastian Mair and Ulf Brefeld. "Coresets for Archetypal Analysis". In: Dec. 2019.

[112] Stefano Mariani, Giacomo Cabri, and Franco Zambonelli. "Coordination of Autonomous Vehicles: Taxonomy and Survey". In: *ACM Computing Surveys* 54.1 (Feb. 2021), 19:1–19:33. ISSN: 0360-0300. DOI: 10.1145/3431231. URL: https://dl.acm.org/doi/10.1145/3431231 (visited on 04/20/2024).

[113] Alejandro G. Martín et al. “Combining user behavioural information at the feature level to enhance continuous authentication systems”. In: *Knowledge-Based Systems* 244 (May 2022), p. 108544. ISSN: 0950-7051. DOI: 10.1016/j.knosys.2022.108544. URL: https://www.sciencedirect.com/science/article/pii/S0950705122002398 (visited on 04/21/2024).

[114] Michael Mateas. “Expressive AI: A Hybrid Art and Science Practice”. In: *Leonardo* 34.2 (2001), pp. 147–153. ISSN: 0024-094X. DOI: 10.1162/002409401750184717.

[115] Catharine L. R. McGhan, Ali Nasir, and Ella M. Atkins. “Human Intent Prediction Using Markov Decision Processes”. In: *Journal of Aerospace Information Systems* 12.5 (2015). Publisher: American Institute of Aeronautics and Astronautics _eprint: https://doi.org/10.2514/1.I010090, pp. 393–397. DOI: 10.2514/1.I010090. URL: https://doi.org/10.2514/1.I010090 (visited on 02/22/2022).

[116] Paul D. McNicholas. “Model-Based Clustering”. en. In: *Journal of Classification* 33.3 (Oct. 2016), pp. 331–373. ISSN: 1432-1343. DOI: 10.1007/s00357-016-9211-9. URL: https://doi.org/10.1007/s00357-016-9211-9 (visited on 04/21/2024).

[117] Paul D. McNicholas and T. Brendan Murphy. “Model-based clustering of longitudinal data”. en. In: *Canadian Journal of Statistics* 38.1 (2010). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cjs.10047, pp. 153–168. ISSN: 1708-945X. DOI: 10.1002/cjs.10047. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/cjs.10047 (visited on 04/21/2024).

[118] Sorvall Microcentrifuge. “Apprenticeship Learning via Inverse Reinfor”. In: 346 (2011). ISBN: 1581138285, pp. 1–2. URL: https://ai.stanford.edu/$%5Csim$ang/papers/icml04-apprentice.pdf.

[119] Vladimir M. Mladenovic, Dana Porrat, and Miroslav D. Lutovac. “The direct execution of the expectation-maximization algorithm using symbolic processing”. In: *2011 10th International Conference on Telecommunication in Modern Satellite Cable and Broadcasting Services (TELSIKS).*

Vol. 1. Oct. 2011, pp. 265–268. DOI: 10.1109/âĎąSKS.2011.6112048. URL: https://ieeexplore.ieee.org/document/6112048 (visited on 04/22/2024).

[120] Volodymyr Mnih et al. "Asynchronous Methods for Deep Reinforcement Learning". en. In: *Proceedings of The 33rd International Conference on Machine Learning*. ISSN: 1938-7228. PMLR, June 2016, pp. 1928–1937. URL: https://proceedings.mlr.press/v48/mniha16.html (visited on 02/26/2022).

[121] Volodymyr Mnih et al. "Playing Atari with Deep Reinforcement Learning". In: *arXiv:1312.5602 [cs]* (Dec. 2013). arXiv: 1312.5602. URL: http://arxiv.org/abs/1312.5602 (visited on 02/26/2022).

[122] Siti Monalisa et al. "Customer segmentation with RFM models and demographic variable using DBSCAN algorithm". en. In: *TELKOMNIKA (Telecommunication Computing Electronics and Control)* 21.4 (Aug. 2023). Number: 4, pp. 742–749. ISSN: 2302-9293. DOI: 10.12928/telkomnika.v21i4.22759. URL: http://telkomnika.uad.ac.id/index.php/TELKOMNIKA/article/view/22759 (visited on 05/11/2024).

[123] Maxim Mozgovoy and Iskander Umarov. "Behavior Capture: Building Believable and Effective AI Agents for Video Games". In: *International Journal of Arts and Sciences* (2011). URL: http://web-ext.u-aizu.ac.jp/$%5Csim$mozgovoy/homepage/papers/mu11b.pdf.

[124] Anusha Nagabandi et al. "Learning to adapt in dynamic, real-world environments through meta-reinforcement learning". In: *arXiv* (2018). ISBN: 9781450371278 _eprint: 1803.11347, pp. 89–90. ISSN: 23318422.

[125] Anusha Nagabandi et al. "Neural Network Dynamics for Model-Based Deep Reinforcement Learning with Model-Free Fine-Tuning". In: *Proceedings - IEEE International Conference on Robotics and Automation* (2018). ISBN: 9781538630815 _eprint: 1708.02596, pp. 7579–7586. ISSN: 10504729. DOI: 10.1109/ICRA.2018.8463189.

[126] Jyoti Narwariya et al. "Meta-Learning for Few-Shot Time Series Classification". In: (). ISBN: 9781450377386, pp. 28–36.

[127] Cyrus Neary et al. "Reward Machines for Cooperative Multi-Agent Reinforcement Learning". In: (2020). _eprint: 2007.01962. URL: http://arxiv.org/abs/2007.01962.

[128] C.L. Nehaniv et al. "A methodological approach relating the classification of gesture to identification of human intent in the context of human-robot interaction". In: *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.* ISSN: 1944-9437. Aug. 2005, pp. 371–377. DOI: 10.1109/ROMAN.2005.1513807.

[129] OpenAI et al. "Dota 2 with Large Scale Deep Reinforcement Learning". In: *arXiv:1912.06680 [cs, stat]* (Dec. 2019). arXiv: 1912.06680. URL: http://arxiv.org/abs/1912.06680 (visited on 02/26/2022).

[130] Jelili Oyelade et al. "Clustering Algorithms: Their Application to Gene Expression Data". en. In: *Bioinformatics and Biology Insights* 10 (Jan. 2016). Publisher: SAGE Publications Ltd STM, BBI.S38316. ISSN: 1177-9322. DOI: 10.4137/BBI.S38316. URL: https://doi.org/10.4137/BBI.S38316 (visited on 04/21/2024).

[131] Niko Pajkovic. "Algorithms and taste-making: Exposing the Netflix Recommender System's operational logics". en. In: *Convergence* 28.1 (Feb. 2022). Publisher: SAGE Publications Ltd, pp. 214–235. ISSN: 1354-8565. DOI: 10.1177/13548565211014464. URL: https://doi.org/10.1177/13548565211014464 (visited on 04/20/2024).

[132] Nikolaos Papachristou et al. "Comparing machine learning clustering with latent class analysis on cancer symptoms' data". In: *2016 IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT).* Nov. 2016, pp. 162–166. DOI: 10.1109/HIC.2016.7797722.

[133] Jack Parker-Holder et al. "Effective Diversity in Population Based Reinforcement Learning". In: *arXiv:2002.00632 [cs, stat]* (Oct. 2020). arXiv:

2002.00632. URL: http://arxiv.org/abs/2002.00632 (visited on 02/26/2022).

[134] Simon Parsons and Michael Wooldridge. “Game Theory and Decision Theory in Multi-Agent Systems”. en. In: *Autonomous Agents and Multi-Agent Systems* 5.3 (Sept. 2002), pp. 243–254. ISSN: 1573-7454. DOI: 10.1023/A:1015575522401. URL: https://doi.org/10.1023/A:1015575522401 (visited on 11/07/2021).

[135] Chris Pelling and Henry Gardner. “Two human-like imitation-learning bots with probabilistic behaviors”. In: *IEEE Conference on Computatonal Intelligence and Games, CIG* 2019-Augus (2019). ISBN: 9781728118840 Publisher: IEEE. ISSN: 23254289. DOI: 10.1109/CIG.2019.8847995.

[136] Mihai Polceanu. “MirrorBot: Using human-inspired mirroring behavior to pass a turing test”. In: *IEEE Conference on Computatonal Intelligence and Games, CIG* (2013). ISBN: 9781467353113 Publisher: IEEE. ISSN: 23254270. DOI: 10.1109/CIG.2013.6633618.

[137] Amy R. Pritchett, So Young Kim, and Karen M. Feigh. “Measuring Human-Automation Function Allocation”. en. In: *Journal of Cognitive Engineering and Decision Making* 8.1 (Mar. 2014), pp. 52–77. ISSN: 1555-3434. DOI: 10.1177/1555343413490166. URL: http://journals.sagepub.com/doi/10.1177/1555343413490166 (visited on 04/21/2024).

[138] Mateusz Przyborowski and Dominik Ślęzak. “Approximation of the expectation-maximization algorithm for Gaussian mixture models on big data”. In: *2022 IEEE International Conference on Big Data (Big Data)*. Dec. 2022, pp. 6256–6260. DOI: 10.1109/BigData55660.2022.10020450. URL: https://ieeexplore.ieee.org/document/10020450 (visited on 04/22/2024).

[139] Teng Qiu and Yong-Jie Li. “Fast LDP-MST: An Efficient Density-Peak-Based Clustering Method for Large-Size Datasets”. en. In: *IEEE Transactions on Knowledge and Data Engineering* 35.5 (May 2023), pp. 4767–4780. ISSN: 1041-4347, 1558-2191, 2326-3865. DOI: 10.1109/TKDE.2022.

3150403. URL: https://ieeexplore.ieee.org/document/9712197/ (visited on 04/16/2024).

[140] Jack W. Rae et al. "Fast Parametric Learning with Activation Memorization". In: *arXiv:1803.10049 [cs, stat]* (Mar. 2018). arXiv: 1803.10049. URL: http://arxiv.org/abs/1803.10049 (visited on 02/26/2022).

[141] David G. Rand. "Cooperation, Fast and Slow: Meta-Analytic Evidence for a Theory of Social Heuristics and Self-Interested Deliberation". en. In: *Psychological Science* 27.9 (Sept. 2016), pp. 1192–1206. ISSN: 0956-7976, 1467-9280. DOI: 10.1177/0956797616654455. URL: http://journals.sagepub.com/doi/10.1177/0956797616654455 (visited on 01/26/2022).

[142] David G. Rand et al. "Social heuristics shape intuitive cooperation". en. In: *Nature Communications* 5.1 (Apr. 2014). Number: 1 Publisher: Nature Publishing Group, p. 3677. ISSN: 2041-1723. DOI: 10.1038/ncomms4677. URL: https://www.nature.com/articles/ncomms4677 (visited on 02/02/2022).

[143] Laura A. Ripamonti et al. "Believable group behaviours for NPCs in FPS games". In: *Proceedings - IEEE Symposium on Computers and Communications* Iscc (2017). ISBN: 9781538616291, pp. 12–17. ISSN: 15301346. DOI: 10.1109/ISCC.2017.8024497.

[144] Casey Rosenthal and Clare Bates Congdon. "Personality profiles for generating believable bot behaviors". In: *2012 IEEE Conference on Computational Intelligence and Games, CIG 2012* (2012). ISBN: 9781467311922 Publisher: IEEE, pp. 124–131. DOI: 10.1109/CIG.2012.6374147.

[145] Maurice Roux. "A Comparative Study of Divisive and Agglomerative Hierarchical Clustering Algorithms". en. In: *Journal of Classification* 35.2 (July 2018), pp. 345–366. ISSN: 1432-1343. DOI: 10.1007/s00357-018-9259-9. URL: https://doi.org/10.1007/s00357-018-9259-9 (visited on 04/22/2024).

[146] Richard M. Ryan, C. Scott Rigby, and Andrew Przybylski. "The Motivational Pull of Video Games: A Self-Determination Theory Approach". en. In: *Motivation and Emotion* 30.4 (Dec. 2006), pp. 344–360. ISSN: 1573-6644. DOI: 10.1007/s11031-006-9051-8. URL: https://doi.org/10.1007/s11031-006-9051-8 (visited on 11/07/2021).

[147] Kawtar Sabor et al. "A data mining approach for improved interpretation of ERT inverted sections using the DBSCAN clustering algorithm". In: *Geophysical Journal International* 225.1 (Dec. 2020). Conference Name: Geophysical Journal International, pp. 1304–1318. ISSN: 1365-246X. DOI: 10.1093/gji/ggab023. URL: https://ieeexplore.ieee.org/document/9434316 (visited on 04/19/2024).

[148] Tim Salimans et al. "Evolution Strategies as a Scalable Alternative to Reinforcement Learning". In: *arXiv:1703.03864 [cs, stat]* (Sept. 2017). arXiv: 1703.03864. URL: http://arxiv.org/abs/1703.03864 (visited on 02/26/2022).

[149] Stefan Schaal et al. "Learning human-like Movement Behavior for Computer Games". In: *From animals to animats 8: Proceedings of the Eighth International Conference on the Simulation of Adaptive Behavior*. Conference Name: From animals to animats 8: Proceedings of the Eighth International Conference on the Simulation of Adaptive Behavior. MIT Press, 2004, pp. 315–323. ISBN: 978-0-262-29144-6. URL: https://ieeexplore.ieee.org/document/6281990 (visited on 05/02/2024).

[150] Mikkel N. Schmidt et al. "Latent profile analysis of human values: What is the optimal number of clusters?" In: *Methodology* 17.2 (June 2021), pp. 127–148. ISSN: 1614-2241. DOI: 10.5964/meth.5479. URL: https://meth.psychopen.eu/index.php/meth/article/view/5479 (visited on 05/02/2024).

[151] Rahma Wati Sembiring Brahmana, Fahd Agodzo Mohammed, and Kankamol Chairuang. "Customer Segmentation Based on RFM Model Using K-Means, K-Medoids, and DBSCAN Methods". en. In: *Lontar Komputer :*

*Jurnal Ilmiah Teknologi Informasi* 11.1 (Apr. 2020), p. 32. ISSN: 2541-5832, 2088-1541. DOI: 10.24843/LKJITI.2020.v11.i01.p04. URL: https://ojs.unud.ac.id/index.php/lontar/article/view/58025 (visited on 05/11/2024).

[152] Sohan Seth and Manuel J. A. Eugster. "Probabilistic archetypal analysis". en. In: *Machine Learning* 102.1 (Jan. 2016), pp. 85–113. ISSN: 0885-6125, 1573-0565. DOI: 10.1007/s10994-015-5498-8. URL: http://link.springer.com/10.1007/s10994-015-5498-8 (visited on 10/13/2021).

[153] Yunxiao Shan et al. "A Density Peaks Clustering Algorithm With Sparse Search and K-d Tree". en. In: *IEEE Access* 10 (2022), pp. 74883–74901. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2022.3190958. URL: https://ieeexplore.ieee.org/document/9829753/ (visited on 04/16/2024).

[154] N. D. Shchepin and A. S. Zagarskikh. "Building behavioral AI using trust and reputation model based on mask model." In: *Procedia Computer Science* 156 (2019). Publisher: Elsevier B.V., pp. 387–394. ISSN: 18770509. DOI: 10.1016/j.procs.2019.08.216. URL: https://doi.org/10.1016/j.procs.2019.08.216.

[155] Andriy Shepitsen et al. "Personalized recommendation in social tagging systems using hierarchical clustering". In: *Proceedings of the 2008 ACM conference on Recommender systems*. RecSys '08. New York, NY, USA: Association for Computing Machinery, Oct. 2008, pp. 259–266. ISBN: 978-1-60558-093-7. DOI: 10.1145/1454008.1454048. URL: https://dl.acm.org/doi/10.1145/1454008.1454048 (visited on 05/10/2024).

[156] P. Shunmugapriya and S. Kanmani. "Optimization of stacking ensemble configurations through Artificial Bee Colony algorithm". en. In: *Swarm and Evolutionary Computation* 12 (Oct. 2013), pp. 24–32. ISSN: 2210-6502. DOI: 10.1016/j.swevo.2013.04.004. URL: https://www.sciencedirect.com/science/article/pii/S2210650213000199 (visited on 02/26/2022).

[157] Maarten Sierhuis et al. "Human-Agent Teamwork and Adjustable Autonomy in Practice". en. In: (), p. 8.

[158] David Silver et al. "Mastering the game of Go without human knowledge". en. In: *Nature* 550.7676 (Oct. 2017). Number: 7676 Publisher: Nature Publishing Group, pp. 354–359. ISSN: 1476-4687. DOI: 10.1038/nature24270. URL: https://www.nature.com/articles/nature24270 (visited on 02/17/2022).

[159] Andrey Simonov, Aleksandr Zagarskikh, and Victor Fedorov. "Applying Behavior characteristics to decision-making process to create believable game AI". In: *Procedia Computer Science* 156 (2019). Publisher: Elsevier B.V., pp. 404–413. ISSN: 18770509. DOI: 10.1016/j.procs.2019.08.222. URL: https://doi.org/10.1016/j.procs.2019.08.222.

[160] Priyanka Sinha et al. "A Hierarchical Clustering Algorithm for Characterizing Social Media Users". In: *Companion Proceedings of the Web Conference 2020*. WWW '20. New York, NY, USA: Association for Computing Machinery, Apr. 2020, pp. 353–362. ISBN: 978-1-4503-7024-0. DOI: 10.1145/3366424.3383296. URL: https://dl.acm.org/doi/10.1145/3366424.3383296 (visited on 05/10/2024).

[161] César Soto-Valero. "A Gaussian mixture clustering model for characterizing football players using the EA Sports' FIFA video game system. [Modelo basado en agrupamiento de mixturas Gaussianas para caracterizar futbolistas utilizando el sistema de videojuegos FIFA de EA Sports]." en. In: *RICYDE. Revista Internacional de Ciencias del Deporte. doi:10.5232/ricyde* 13.49 (Mar. 2017). Number: 49, pp. 244–259. ISSN: 1885-3137. URL: https://www.cafyd.com/REVISTA/ojs/index.php/ricyde/article/view/1165 (visited on 05/11/2024).

[162] D. J. Strouse et al. "Collaborating with Humans without Human Data". In: *arXiv:2110.08176 [cs]* (Oct. 2021). arXiv: 2110.08176. URL: http://arxiv.org/abs/2110.08176 (visited on 11/11/2021).

[163] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: A Bradford Book, Oct. 2018. ISBN: 978-0-262-03924-6.

[164] Maciej Swiechowski and Dominik Slezak. "Grail: A Framework for Adaptive and Believable AI in Video Games". In: *Proceedings - 2018 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2018* (2019). ISBN: 9781538673256 Publisher: IEEE, pp. 762–765. DOI: 10.1109/WI.2018.00012.

[165] M. A. Syakur et al. "Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster". en. In: *IOP Conference Series: Materials Science and Engineering* 336.1 (Apr. 2018). Publisher: IOP Publishing, p. 012017. ISSN: 1757-899X. DOI: 10.1088/1757-899X/336/1/012017. URL: https://dx.doi.org/10.1088/1757-899X/336/1/012017 (visited on 05/11/2024).

[166] Panagiotis Symeonidis and Nikolaos Mantas. "Spectral clustering for link prediction in social networks with positive and negative links". en. In: *Social Network Analysis and Mining* 3.4 (Dec. 2013), pp. 1433–1447. ISSN: 1869-5469. DOI: 10.1007/s13278-013-0128-6. URL: https://doi.org/10.1007/s13278-013-0128-6 (visited on 04/19/2024).

[167] Ruck Thawonmas and Keita Iizuka. "Visualization of Online-Game Players Based on Their Action Behaviors". In: *International Journal of Computer Games Technology* 2008 (2008), pp. 1–9. ISSN: 1687-7047. DOI: 10.1155/2008/906931.

[168] Christian Thurau and A Drachen. "Introducing Archetypal Analysis for Player Classification in Games Categories and Subject Descriptors". In: *Proceedings of the EPEX 11 Workshop* (2011). ISBN: 9781450308045.

[169] Christian Thurau, Tobias Paczian, and C. Bauckhage. "Is Bayesian Imitation Learning the Route to Believable Gamebots". In: 2004. URL: https://www.semanticscholar.org/paper/Is-Bayesian-Imitation-

Learning-the-Route-to-Thurau-Paczian/f1f888747609a4789fa59f4d92bf063213dba4d5 (visited on 05/02/2024).

[170] Gustavo F. Tondello and Lennart E. Nacke. “Player Characteristics and Video Game Preferences”. In: *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. CHI PLAY ’19. New York, NY, USA: Association for Computing Machinery, Oct. 2019, pp. 365–378. ISBN: 978-1-4503-6688-5. DOI: 10.1145/3311350.3347185. URL: https://dl.acm.org/doi/10.1145/3311350.3347185 (visited on 04/19/2024).

[171] Raimo Tuomela. “What is cooperation?” en. In: *Erkenntnis* 38.1 (Jan. 1993), pp. 87–101. ISSN: 0165-0106, 1572-8420. DOI: 10.1007/BF01129023. URL: http://link.springer.com/10.1007/BF01129023 (visited on 11/06/2021).

[172] Anna-Sophie Ulfert et al. “Shaping a multidisciplinary understanding of team trust in human-AI teams: a theoretical framework”. In: *European Journal of Work and Organizational Psychology* 0.0 (2023). Publisher: Routledge _eprint: https://doi.org/10.1080/1359432X.2023.2200172, pp. 1–14. ISSN: 1359-432X. DOI: 10.1080/1359432X.2023.2200172. URL: https://doi.org/10.1080/1359432X.2023.2200172 (visited on 04/19/2024).

[173] Karel Van Den Bosch and Adelbert Bronkhorst. “Human-AI Cooperation to Benefit Military Decision Making”. In: *Sto* June (2018), pp. 1–12. URL: https://www.researchgate.net/profile/Karel_Bosch/publication/325718292_Human-AI_Cooperation_to_Benefit_Military_Decision_Making/links/5b1fc0b2a6fdcc69745cd169/Human-AI-Cooperation-to-Benefit-Military-Decision-Making.pdf.

[174] Jakob Veselsky et al. “Establishing trust in vehicle-to-vehicle coordination: a sensor fusion approach”. In: *Proceedings of the 23rd Annual International Workshop on Mobile Computing Systems and Applications*. HotMobile ’22. New York, NY, USA: Association for Computing Machin-

ery, Mar. 2022, p. 128. ISBN: 978-1-4503-9218-1. DOI: 10.1145/3508396.3517075. URL: https://dl.acm.org/doi/10.1145/3508396.3517075 (visited on 04/19/2024).

[175] Alexander Sasha Vezhnevets et al. “FeUdal Networks for Hierarchical Reinforcement Learning”. In: *arXiv:1703.01161 [cs]* (Mar. 2017). arXiv: 1703.01161. URL: http://arxiv.org/abs/1703.01161 (visited on 02/26/2022).

[176] Jennifer Villareale and Jichen Zhu. “Understanding Mental Models of AI through Player-AI Interaction”. In: (2021). _eprint: 2103.16168. URL: http://arxiv.org/abs/2103.16168.

[177] Jue Wang et al. “Clustering Analysis of Human Behavior Based on Mobile Phone Sensor Data”. In: *Proceedings of the 2018 10th International Conference on Machine Learning and Computing*. ICMLC ’18. New York, NY, USA: Association for Computing Machinery, Feb. 2018, pp. 64–68. ISBN: 978-1-4503-6353-2. DOI: 10.1145/3195106.3195146. URL: https://dl.acm.org/doi/10.1145/3195106.3195146 (visited on 05/01/2024).

[178] Lijuan Wang, Shifei Ding, and Hongjie Jia. “An Improvement of Spectral Clustering via Message Passing and Density Sensitive Similarity”. en. In: *IEEE Access* 7 (2019), pp. 101054–101062. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2019.2929948. URL: https://ieeexplore.ieee.org/document/8766800/ (visited on 04/19/2024).

[179] Pei-pei Wang et al. “A recommendation algorithm based on density peak clustering and key users”. In: *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*. July 2017, pp. 468–473. DOI: 10.1109/FSKD.2017.8393314. URL: https://ieeexplore.ieee.org/abstract/document/8393314 (visited on 05/11/2024).

[180] David Watson, Ericka Nus, and Kevin D. Wu. “Development and Validation of the Faceted Inventory of the Five-Factor Model (FI-FFM)”. eng. In: *Assessment* 26.1 (Jan. 2019), pp. 17–44. ISSN: 1552-3489. DOI: 10.1177/1073191117711022.

[181] Thomas A. Widiger and Jennifer Ruth Presnall. “Clinical Application of the Five-Factor Model”. en. In: *Journal of Personality* 81.6 (2013). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jopy.12004, pp. 515–527. ISSN: 1467-6494. DOI: 10.1111/jopy.12004. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jopy.12004 (visited on 04/21/2024).

[182] Tengke Xiong et al. “DHCC: Divisive hierarchical clustering of categorical data”. en. In: *Data Mining and Knowledge Discovery* 24.1 (Jan. 2012), pp. 103–135. ISSN: 1573-756X. DOI: 10.1007/s10618-011-0221-2. URL: https://doi.org/10.1007/s10618-011-0221-2 (visited on 04/22/2024).

[183] Nacim Yanes et al. “A Machine Learning-Based Recommender System for Improving Students Learning Experiences”. In: *IEEE Access* 8 (2020). Conference Name: IEEE Access, pp. 201218–201235. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2020.3036336. URL: https://ieeexplore.ieee.org/document/9249379 (visited on 04/20/2024).

[184] Zhao Yilin, Amjad Fayoumi, and Azar Shahgholian. “Understanding Online Customer Touchpoints: A Deep Learning Approach to Enhancing Customer Experience in Digital Retail”. In: *2023 9th International Conference on Information Technology Trends (ITT)*. May 2023, pp. 193–198. DOI: 10.1109/ITT59889.2023.10184269. URL: https://ieeexplore.ieee.org/document/10184269 (visited on 06/02/2024).

[185] Lifeng Yin et al. “An Improved Density Peak Clustering Algorithm for Multi-Density Data”. en. In: *Sensors* 22.22 (Jan. 2022). Number: 22 Publisher: Multidisciplinary Digital Publishing Institute, p. 8814. ISSN: 1424-8220. DOI: 10.3390/s22228814. URL: https://www.mdpi.com/1424-8220/22/22/8814 (visited on 04/16/2024).

[186] Xiaohong Yu et al. "Trajectory dimensionality reduction and hyperparameter settings of DBSCAN for trajectory clustering". en. In: *IET Intelligent Transport Systems* 16.5 (2022). ISSN: 1751-9578. DOI: 10.1049/itr2.12166. URL: https://onlinelibrary.wiley.com/doi/abs/10.1049/itr2.12166 (visited on 04/19/2024).

[187] Huaqing Zhang and Shanghang Zhang. "Multi-agent reinforcement learning". In: *Deep Reinforcement Learning: Fundamentals, Research and Applications* 73 (2020). ISBN: 9789811540950, pp. 335–346. DOI: 10.1007/978-981-15-4095-0_11.

[188] Yu Zhang. "Role-based shared mental models". In: *2008 International Symposium on Collaborative Technologies and Systems*. May 2008, pp. 424–431. DOI: 10.1109/CTS.2008.4543960. URL: https://ieeexplore.ieee.org/document/4543960 (visited on 04/21/2024).

[189] Ruiqi Zheng et al. "AutoML for Deep Recommender Systems: A Survey". In: *ACM Transactions on Information Systems* 41.4 (Mar. 2023), 101:1–101:38. ISSN: 1046-8188. DOI: 10.1145/3579355. URL: https://dl.acm.org/doi/10.1145/3579355 (visited on 04/20/2024).

[190] Xianrui Zheng, Chao Zhang, and Philip C. Woodland. "Adapting GPT, GPT-2 and BERT Language Models for Speech Recognition". In: *2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. Dec. 2021, pp. 162–168. DOI: 10.1109/ASRU51503.2021.9688232. URL: https://ieeexplore.ieee.org/document/9688232 (visited on 06/02/2024).

[191] Jiale Zhi et al. "Fiber: A Platform for Efficient Development and Distributed Training for Reinforcement Learning and Population-Based Methods". In: *arXiv:2003.11164 [cs, stat]* (Mar. 2020). arXiv: 2003.11164. URL: http://arxiv.org/abs/2003.11164 (visited on 02/26/2022).

[192] Fan Zhou et al. "Meta-GNN: On few-shot node classification in graph meta-learning". In: *arXiv* (2019). ISBN: 9781450369763, pp. 2357–2360. ISSN: 23318422.