Acta Ludologica

2024, *Vol. 7, No. 2*

Smart NPCs with Personality in a Serious Game Using Machine Learning

Georgios Liapis, Ioannis Vlahavas

MSc. Georgios Liapis
Aristotle University of Thessaloniki
School of Informatics
Aristotle University Campus
541 24 Thessaloniki
GREECE
gliapisa@csd.auth.gr

Georgios Liapis received his diploma in Applied Informatics (2019) from University of Macedonia, Thessaloniki Greece, and his master's degree in Artificial Intelligence (2021) from Aristotle University of Thessaloniki. Since 2021, he has been a PhD candidate in the Department of Informatics, AUTH, under the supervision of Prof. Ioannis Vlahavas. His research interests are deep reinforcement learning, machine learning, and AI in serious games. He has successfully presented his work in national and international conferences and takes part in summer schools like the AI and Games Summer School 2022 (received full scholarship from Sony AI) and the Mediterranean Machine Learning Summer School 2023.

prof. Ioannis Vlahavas, Ph.D.
Aristotle University of Thessaloniki
School of Informatics
Aristotle University Campus
541 24 Thessaloniki
GREECE
vlahavas@csd.auth.gr

Ioannis Vlahavas is a professor at the School of Informatics at the Aristotle University of Thessaloniki. He has been a visiting scholar at the Department of CS at Purdue University and in 2017 was elected EurAI Fellow from the European Association for Artificial Intelligence. He specializes in knowledge based and AI (machine learning) systems and he has published over 350 papers and book chapters, and co-authored 9 books in these areas. Google Scholar gives a number of 19,066 citations and an h-index of 58. He teaches AI, machine learning, and DSS. He was the chairman of numerous conferences and leads the Intelligent Systems Lab in the department of informatics at Aristotle University of Thessaloniki.

ABSTRACT:
Gaming technology's potential extends beyond entertainment, providing a powerful platform for learning and evaluation, and for that, NPCs with static movement and conversation behaviours are often used. To make them more human-like and emulate actions, technologies such as artificial intelligence are utilized. This work proposes smart NPCs to imitate personality traits in a serious escape room setting. For their development, labelled personality profiles are normally required from human players to define their standard behaviours. As this is rather difficult, deep reinforcement learning is a feasible and effective alternative for generating the necessary dataset. Each NPC is an AI agent that simulates a specific personality according to the OCEAN 5 model. Our escape room environment also includes Raven-inspired intelligence tests and a custom communication system that allows the development of smart NPC teams. Analysis of gameplay data and metrics uncovered behavioural patterns affecting performance, stability, and task completion times. Such progress has potential across multiple digital game types for smart NPCs with specific personality, as well as for the creation of standard gameplay style profiles that can be used for players' assessment.

KEY WORDS:
agents, deep reinforcement learning, machine learning, NPC, serious games.

DOI:
10.34135/actaludologica.2024-7-2.4-25

Introduction

As technology advances rapidly, it opens new opportunities and pushes numerous scientific fields to unprecedented new heights. One such area undergoing deep transition is gaming, which has evolved beyond simply amusement. Gaming aspects are easily integrated into approaches in various disciplines, including human resource management and education, to increase efficacy and widen vistas.

This paradigm shift is encapsulated within gamification-based systems, a concept harnessing gaming characteristics to influence behaviours within non-game contexts (Robson et al., 2015). Central to gamified systems are rewards, challenges, profiling, and leaderboards, seamlessly woven into numerous sectors spanning education, businesses, and marketing endeavours (Dicheva et al., 2015).

This approach heralds significant potential in another area, for the assessment of individuals across diverse domains, as their profiles naturally emerge through immersive gameplay experiences. Leveraging gamification techniques, organizations can glean nuanced insights into individuals' capabilities, fostering more informed decision-making processes and catalysing ongoing improvement endeavours. The overarching goal of gamification is multifaceted, aiming not only to enhance player skills and critical thinking, but also to yield valuable metrics and insights into gameplay styles and profiles within the context of serious gaming (Abt, 1987). There are games that are designed for purposes other than only entertainment.

One such case are *escape room* (ER) games, whether in physical or virtual formats. ERs challenge players or teams to surmount physical and mental obstacles within a constrained timeframe to secure their escape. Given the paramount importance of effective teamwork and communication for success, companies increasingly turn to ERs for team-building exercises and comprehensive assessments of individual and collective performance.

However, conventional methods employed in real-life ERs to gauge team or personal dynamics and individual effectiveness often fall short, relying predominantly on post-room questionnaires susceptible to biases and limitations. Monitoring each player's movements, actions, and interactions throughout an entire ER session, typically lasting an hour for teams comprising 4 to 7 members, presents logistical challenges (Fotaris & Mastoras, 2019).

In this work an innovative solution is developed in the form of *MindEscape*, a serious ER digital game that not only provides an immersive gaming experience but also a comprehensive simulation environment. *MindEscape*, created with the Unity game engine, incorporates intelligence tests and a custom communication system, allowing for in-depth monitoring of player interactions. *MindEscape* encapsules the main aspects of serious games in the form of ER and can assist in many gamified processes with its results and generated data, like the assessment of player profiles.

It must be noted that a vital part of games are NPCs, performing a variety of duties such as delivering missions, offering aid, acting as foes or friends, or just adding dimension to the game environment. Most of the times NPCs can be written characters with predefined behaviours and conversations. They add to the immersion and storyline of the game, making the virtual environment feel more alive and dynamic. Additionally, AI can produce characters that are more dynamic and human like.

Using *deep reinforcement learning* (DRL) agents, a revolutionary tool was developed capable of imitating template behaviours based on the OCEAN 5 personality model, subject to predefined reward functions. Through recurrent training cycles across multiple scenarios and behavioural profiles, these agents create abundant labelled gaming data, eliminating the need for considerable human-player engagement. This results in the formation of agents who may function as NPCs in the ER scenario, each with their own personality.

Furthermore, the work of Durupinar et al. (2011) is a foundation for this work by using the default behaviours generated within simulated environments to validate and contextualize the trained agents' gameplay styles. This ensures fidelity in correlating simulated behaviours with emulated personality traits, providing valuable insights into individual intelligence scores and behavioural tendencies.

These final NPCs with specific personality traits can be evaluated in different kinds of scenarios and game types, so that results may be drawn of how behaviours play a crucial role in game analysis. Moreover, the custom reward functions can be also utilized as a basis for a new way to develop smart NPCs in any digital game environment.

The ramifications of our findings extend beyond mere player assessments, offering insights into group dynamics and collaborative problem-solving approaches across diverse demographics, from children to corporate teams. By elucidating commonalities and disparities in how individuals and teams navigate shared objectives and diverse scenarios within a controlled environment, our study lays the groundwork for an ethical and reliable assessment framework ("Ethics guidelines for trustworthy AI", 2019). The contributions of this work are new methodologies for:

- the design and implementation of NPC in the form of DRL agents, each emulating a personality trait.
- encapsulating personality in a reward system for DRL agents that can be used in other dynamic environments and games.
- developing intelligent NPCs that can solve IQ tests.
- evaluation of teams of agents' efficiency based on their personality.

Related Work

The role of ERs in fostering team cohesion and serving as a gamification tool across various scientific domains has garnered significant research attention. Clarke et al. (2017), for instance, highlight the utility of ERs in educational settings, showcasing their effectiveness in cultivating soft skills through puzzle-solving experiences. In their study, participants engaged in a real-world pilot room, tasked with defusing a bomb within a 15-minute timeframe. Post-game, teams provided feedback on the educational value of the experience via surveys. A distinguishing aspect of our work lies in the creation of a simulated ER environment, providing a virtual platform (i.e. a video game) to compute and assess multiple evaluation metrics, thereby offering deeper insights into each player's gameplay style and overall performance. Moreover, the integration of DRL agents enables the simulation of diverse gameplay styles and behaviours, enriching the dataset compared to real-life player data collection.

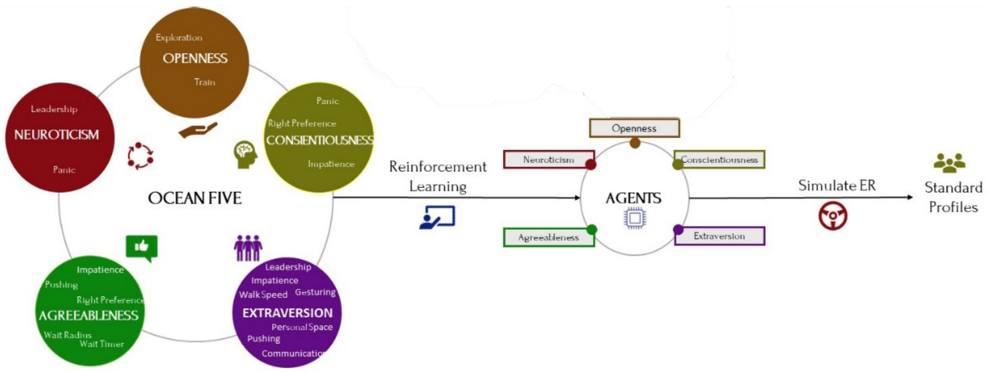
Furthermore, online ERs have also been leveraged in educational contexts, as demonstrated by Vergne et al. (2020). Utilizing online document editors, presentation software, and video conferencing tools, their game served as a remote learning method during 2020. Players navigated various challenges presented as rooms within a factory-themed environment, communicating via video to solve problems within a 20-minute timeframe. While this online game facilitated direct communication, it lacked the depth of 3D exploration and interaction found in *MindEscape*. Additionally, the implementation of DRL agents in *MindEscape* yields a rich array of diverse gameplay data and outcomes, enabling a more comprehensive evaluation of the game's efficacy. In addition, prior efforts are acknowledged in utilizing game environments to model behaviours. Liapis et al. (2021) explored how a single agent could emulate basic in-game behaviours, focusing on movement influenced by the agent's openness personality trait within a simplistic room setting. Similarly, another study proposed the theoretical use of an ER to capture specific gameplay data and metrics indicative of a player's personality through tailored puzzles and riddles (see Liapis et al., 2022).

In contrast, this work extends beyond these approaches by implementing a single and team agent system. A reward function methodology was introduced tailored for teams, facilitating the measurement of efficiency within a gamified environment. By incorporating complex behaviours influenced by multiple personality traits, our approach provides a more comprehensive assessment of individual and team dynamics within the game setting.

Lastly, research has also explored personality trait assessment methodologies. While self-evaluation questionnaires remain prevalent, they are susceptible to biases and linguistic variations. Notably, models such as the OCEAN 5 and Personality Inventory and their revisions are widely used, and being assessed with questionnaires like the NEO PI-R (Costa & McCrae, 2008) and TPQue (Tsaousis, 1998). These tests offer insights into individuals' personality traits, facilitating a more holistic understanding for assessment purposes.

Methodology

In this section, the main aspects of our game are presented, i.e. the intelligence tests and the personality traits model while also delving into the intricate gameplay mechanics of *MindEscape* as a playground for the NPC agents, detailing the framework for administering IQ tests, defining metrics for measuring personality traits and analysing the implementation of DRL agents. The workflow of the paper and how the final standard profiles are generated is graphically represented in Picture 1.

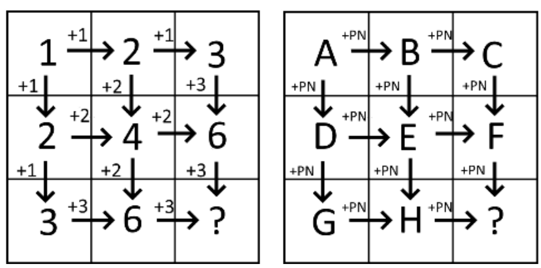


Picture 1: System workflow: The OCEAN 5 Personality traits are modelled as mathematical equations and are used as ground truth for the reinforcement learning agents' rewards systems

Source: own processing

a) Background

The intelligence tests embedded within the gameplay draw inspiration from the renowned Raven Intelligence tests (Raven, 2000), resembling interactive mini games. These tests mirror the structure commonly found in traditional IQ assessments, featuring grids comprising nine shapes arranged in patterns across rows and columns. As depicted in Picture 2, exemplifying an arithmetic Raven IQ test, the numerical values ascend by increments of 1, 2, or 3 within each row or column, culminating in the solution represented by the number 9.



Picture 2: Raven IQ test example and the methodology it follows

Source: own processing

In *MindEscape*, the OCEAN 5 (Jang et al., 1996) personality traits model is integrated. This acronym encompasses Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, constituting one of the most established frameworks in the field.

Personality assessment is a multifaceted domain, encompassing traits such as temperament, emotional disposition, and cognitive tendencies, among others. The precise enumeration of these traits remains a subject of considerable debate, prompting an extensive investigation by numerous researchers. The OCEAN 5 model emerged from this body of research, synthesizing key dimensions for comprehensive personality evaluation.

While further modifications and alternative models have emerged, such as the Psychopathic Personality Inventory (Uzieblo et al., 2010) and its iterations, they share fundamental concepts with the OCEAN 5 framework. These newer models often refine and restructure subcategories within each feature. However, for *MindEscape*, OCEAN 5 model was chosen because of its extensive adoption, universal acceptance, and broad application across multiple situations.

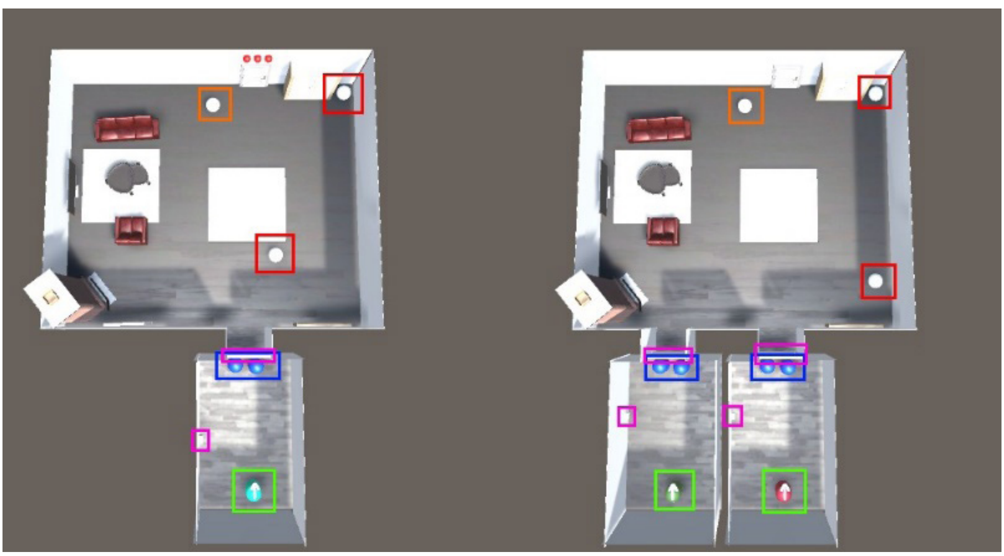
Reinforcement learning stands as a pivotal domain within machine learning, centred on training agents to navigate and learn from interactions within unfamiliar environments. Typically, these agents strive to select actions that maximize a predefined reward function within a given scenario or environment, leveraging past experiences to inform their decision-making process. Through iterative feedback loops, agents learn to refine their strategies to optimize reward acquisition, thus evolving their capabilities through self-learning (Goldman, 2020).

In our endeavour, agents facilitated by the Unity Machine Learning Agents (ML-Agents) developed a package, utilizing the Proximal Policy Optimization (PPO) algorithm (see Schulman et al., 2017). In subsequent chapters, the methodology was employed in our research as well as the related works in the realm of serious ER games, elucidating their respective objectives. Finally, a comprehensive analysis of the outcomes derived from agent training is analysed and pertinent conclusions are drawn from our findings.

b) Implementation

MindEscape is structured around several rooms, each meticulously themed (e.g. an office), and adorned with furniture and assets. These rooms are further categorized into five levels of difficulty, contingent upon factors such as the intricacy of individual rooms, furniture arrangements, and test complexities.

Within each room, players embark on a mission to discover a set number of test panels, typically four in number. Upon interaction with each panel, players are presented with an IQ test, the successful completion of which enables progression within the room. To escape a room, players must locate and solve all four tests. To accomplish this, they must uncover the initial two or three (depending on difficulty) visible puzzle panels and solve the respective tests. Any hidden panels necessitate discovery via uncovering buttons or inputting hidden numbers on a keyboard scattered throughout the room. Upon resolving all preceding tests, the final one is unveiled, signalling the culmination of the room's challenges.



Picture 3: Single and agents' team environment. Agents are marked by the green box, default NPCs by the blue box, buttons and doors by the purple box, and goals in red (the visible ones) and orange (the hidden ones).
Source: own processing

Finally, two separate modes have been developed: a single-agent and an agent's team environment, both following a similar structure but with subtle variances, as shown in this section. Picture 3 shows graphic depictions of these gaming worlds. Here, the agent(s) can be seen, default NPCs/bots, a button, and the opened door. Additionally, three objectives or goals are visible, signifying the targets the agent must meet. These chambers are designed using simplified escape-the-room game logic.

It must be noted that the second environment is not multi-agent, but rather a team consisting of individually trained agents since there is no team reward or team training involved, so that the behaviour of pre-trained agents in a different environment can be examined, where they try to cooperate.

Initially, the agent finds themself confined within the starting room, where they must locate a dynamically positioned button in each playthrough. Upon discovering the button, the agent must either await the departure of default NPCs or manipulate their movements to gain access to the main room. Within the main room, the agent must seek out two goals, also situated in varying locations, before uncovering the final hidden goal and executing their escape. It is worth noting that in the agents' team environment, expansion possibilities include incorporating multiple starting rooms or introducing additional agents into the existing setup.

The primary distinction between the two environments lies in the mechanics surrounding the button-door interaction. In the agents' team environment, each button unlocks a door in the adjacent room, fostering collaborative gameplay dynamics. Moreover, the number of goals remains independent of the number of players, serving as a fundamental gameplay feature consistent across all ER, whether physical or virtual.

Given our focus on developing a serious game, a suite of metrics was meticulously devised, collected, and analysed. The subsequent session delves into the methodology utilized to define and extrapolate these personality metrics.

The IQ test environment was created using the Python programming language, while the tests were created by random numbers depending on the difficulty of the puzzle. To solve these tests, agents must first identify and engage with certain objects (panels) in the room, either via exploration or by solving other riddles that reveal the corresponding panel.

Next, an example based on Picture 1 is presented, where variables from A to H represent numerical values, each following a discernible pattern. Specifically, variables A, B, and C adhere to a predetermined mathematical progression, exemplified by B = A + 1 and C = B + 1 (or equivalently A + 2). Similarly, variables D, E, and F exhibit a parallel pattern, as do G, H, and (?). Hence, it can be inferred that a consistent pattern is applied horizontally across the puzzle.

The player is tasked with identifying the element (?) of the puzzle from a set of six choices. Notably, the starting numbers (A, D, G) always fall within the range of 0 to 50, while the pattern for addition varies between 0 and 10, and for multiplication between 2 and 4. The puzzles are stratified into five difficulty levels: easy (E), easy to medium (EM), medium (M), medium to hard (MH), and hard (H), each characterized by distinct rules as delineated in Table 1.

A key distinction in the Hard level difficulty tests is the introduction of contrasting patterns, rendering them more challenging. Specifically, in this category, the second number (B) is twice the value of the first (B = A * 2), while the third (C) is the result of multiplying the first two numbers (C = A * B). Additionally, it is customary in Raven figure puzzles for the pattern to be applied not only horizontally but also vertically. Adhering to this rule, B = D, C = G, and H = F. Furthermore, alongside the generation of the variables A to H, six possible answers are generated for each test, with one being correct.

Table 1: IQ tests difficulty levels

Difficulty	Operation	First Number	Pattern Number (PN)	Example
E	+	*<= 50*	*<10*	*B = A+PN, C =A+PN*
M	+	*> 50*	*<10*	*B = A+PN, C =B+PN*
M	*	*<= 50*	*<4*	*B = A*PN, C =B*PN*
MH	*	*> 50*	*<4*	*B = A*PN, C =B*PN*
H	*	*> 50*	-	*B=A*2, C = A*B*

Source: own processing

c) Personality Agent Creation Methodology

To obviate the necessity for diverse gameplay data from human players exhibiting varying behavioural characteristics, DRL agents are introduced. Initially, these agents were tasked with emulating the personality traits model, albeit after mastering the Escape room's mechanics through training. Subsequently, metrics from the High-Density Autonomous Crowds system (HiDAC) (Durupinar et al., 2011) were leveraged to adjust rewards for the agents, a process elucidated in subsequent paragraphs. HiDAC serves as a sophisticated crowd simulation system, adept at modelling local behaviours and pathfinding within dynamically evolving environments. An agent's personality π is encapsulated within a five-dimensional vector, with each dimension representing a distinct personality factor, Ψi. The distribution of these personality factors across individuals is modelled via a Gaussian distribution function, characterized by mean (μi) and standard deviation (σi) parameters:

$$\pi = \langle \Psi O, \Psi C, \Psi E, \Psi A, \Psi N \rangle \quad (1)$$

$$\Psi i = (\mu i, \sigma 2) \text{ for } i \in \{O, C, E, A, N\} \text{ where } \mu \in [0, 1], \sigma \in [-0.1, 0.1] \quad (2)$$

The overall behavior β for an individual is a function of different behaviours and is defined as:

$$\beta = (\beta 1, \beta 2, \ldots, \beta n) \text{ , where } \beta j = f(\pi) \text{ , for } j = 1, \ldots, n \quad (3)$$

Given the dynamic nature of each personality trait, the dimension Ψi encompasses a spectrum of values, spanning from positive to negative. Furthermore, behaviours may manifest across multiple personality dimensions, exhibiting varying positive or negative impacts (Table 2). For instance, Leadership might predominantly manifest as a positive trait in individuals exhibiting positive conscientiousness. Conversely, it may manifest as a negative attribute in individuals demonstrating negative agreeableness traits. Alternatively, in individuals with neurotic tendencies, leadership behaviour may simultaneously yield both positive and negative influences.

In Appendix A, mathematical equations delineating each behaviour are provided, derived through meticulous analysis and drawing upon the formulas employed by Durupinar et al. (2011). Importantly, these equations are formulated under the assumption of independence among behaviours. For instance, the behaviour βtr (where "tr" denotes trained) is defined as 1 if the value of ΨO (representing Openness) exceeds or equals 0.5 in the agent's distribution.

Subsequently, the actions and characteristics of the agents are defined, allowing them to exhibit a spectrum of behaviours (refer to Table 3). Subsequent to this, the focus was the practical implications of each behaviour, forging connections with the agents' actions and characteristics. Previous research has provided indications that agents are indeed capable of emulating personality traits within gamified environments (Liapis et al., 2021).

Table 2: Impact of behaviours on personality traits (positive and negative behaviour impact of behaviours to the responding traits)

Behaviour \Traits	**O**		**C**		**E**		**A**		**N**	
	+	-	+	-	+	-	+	-	+	-
Leadership			*X*		*X*	*X*		*X*	*X*	*X*
Trained	*X*	*X*								
Communication					*X*					
Panic			*X*						*X*	*X*
Impatience			*X*		*X*		*X*	*X*		
Pushing			*X*	*X*			*X*	*X*		
Right Preference			*X*	*X*			*X*	*X*		
Avoidance/personal space					*X*	*X*				
Waiting Radius							*X*	*X*		
Waiter Timer								*XX*		
Exploring environment	*X*	*X*								
Walking speed					*X*	*X*				

Source: own processing

State Space

In our Unity implementation, agents leverage a sophisticated observation mechanism centred around Ray-cast observations. This advanced technique harnesses the capabilities of physics functions to cast a ray into the environment scene, providing agents with vital insights upon successful intersection with a target object.

Before embarking on any decision-making process, the agent class diligently invokes this method, enabling agents to gather pertinent information about their surroundings. By harnessing this observation vector, agents can effectively assess their environment, empowering them to make informed choices and navigate through the virtual world with precision and efficiency.

In our system, agents are tailored to consider all relevant information from their environment to make informed decisions. This encompasses details such as the location of buttons, keys, and doors, along with their current state, whether they are pressed, found, or unlocked. Moreover, when one agent interacts with an object, such as picking up a key or pushing a button, this information is promptly relayed to all other agents through their observations.

To capture this comprehensive information, ray-cast arrays are employed projecting a total of 15 rays. Each ray checks for the presence of the 5 specified tags and the 3 vectors representing the environment state. Consequently, the final state space size amounts to 78, encapsulating all pertinent details about the environment essential for the agents' awareness.

Interpreting this wealth of information involves utilizing Boolean values for each tag and vector. This enables the agents to extract knowledge from their sensors and environment, facilitating a deeper understanding of their surroundings and enabling them to make more informed decisions. Throughout the training process, rewards were assigned based on interactions with each element of the environment, as outlined in the subsequent section.

Agent Actions

The agents' actions are dynamically determined during gameplay, while their underlying mechanics are preconfigured based on the personality trait being emulated. For clarity, Table 3 showcases the behaviours as defined by Durupinar et al. (2011) with the corresponding actions exhibited by our agents.

Table 3: Actions to information/description relations

Agent Actions	Information/Description
Walk speed	*3 speeds (step, walk, run)*
Communication system	*"Y" button for Yes*
Indication system	*"N" button for No*
Push actions	*"Q" button to indicate*
Movement information	*"E" button*
Previous knowledge	*Directions angle in each step*
Agent characteristics	**Information/Description**
Collider size	*x, x1,5. X2 scale*
Waiting time	*1,3 or 5 seconds*

Source: own processing

The subsequent phase involves utilizing the accumulated insights, including equations, definitions, etc., to delineate rewards within the game and establish the characteristics of the agents, contingent upon the specific personality trait being assessed (refer to Table 4). Fundamentally, the agents, guided by their actions (outlined in Table 3), generate gameplay data and metrics. Post-completion of each room or episode, rewards are allocated to the agents based on the corresponding trait under evaluation.

Table 4: Behaviours to game mechanics

Original Behaviour	Clarification	Our Actions/Mechanics
Leadership	*Extraversion and stability*	*Movement info*
Trained	*Previous knowledge*	*Knowledge of the map and key element positions*
Communication	*Communication between the team*	*Communication system*
Panic	*Increased walk speed and not waiting*	*Run and pushing*
Impatience	*Route change*	*Running*

Pushing	*Use force to clear the way*	*Use of push button*
Right preference	*Avoiding something from the right side*	*Movement info*
Personal space	*Comfortable territory*	*Collider size*
Wait radius	*Available space needed to move*	*Collider size with need of no collision*
Wait timer	*Wait time in queue*	*Wait time*
Explore	*Numerous actions and increased exploring time*	*Number of actions*
Walk speed	*Movement speed*	*Walk speed*
Gesture	*Nonverbal communication*	*Use of indication*

Source: own processing

Rewards

The following stage involves utilizing the gathered information, including equations, definitions, etc., to delineate the rewards within the game and the attributes of the agents, contingent upon the personality trait under examination (see Table 5). More precisely, the agent, via their actions (see Table 3), generates gameplay data and metrics. Upon the completion of each room episode, the agent receives a reward corresponding to the trait they are being trained on.

A key difference from the equations presented in Appendix A is our amalgamation of calculations for each trait's behaviours into a consolidated framework. Furthermore, the assumption was made that each behaviour exerts an equal percentage of influence on its associated trait. For instance, traits like Leadership and Panic contribute equally, each having a 50% (or 0.5) influence on Neuroticism.

Table 5: Traits to rewards, based on actions and characteristics

Personality Trait	Behaviours (Original)	Reward (custom)	Agent Characteristic
Openness	*Train*	-	*Knowledge of goal positions*
	Explore	*num of correct actions*10*	
Conscientiousness	*Panic*	$0.3 * -2 * \Psi C + 2$ *if run & push*	
	Impatience	$0.3 * (1 - \Psi C)$	
	Right Preference	*If* $\Psi C > 0$ *then* ΨC * *(times right/time)* * 0.3	
Extraversion	*Leadership*	0.3 * *mean speed* * ΨE	
	Communication	*1 if num of communication actions used* $>= \Psi E >= 0.5$	
	Impatience	$0.3 * 2 * \Psi E - 1$ *if* $\Psi E > 0$	
	Pushing	*1 if num of push actions used* $>= 0.3 * \Psi E >= 0.5$	
	Personal Space		*Collider Size*
	Walk speed	*Max walk speed+1*	
	Gesture	*Num of correct gestures * 10*	

Agreeableness	Impatience	0.3 * (1 – ΨA) if run each step	
	Pushing	1 if num of push actions used >= 0.3 * (1- ΨA) >= 0.5	
	Right Preference	0.3 *(Times right/time) * ΨA	
	Wait Radius		Collider Size
	Wait Timer		Wait timer
Neuroticism	Leadership	Mean speed * (1 - ΨN)*0.5	
	Panic	ΨN * 0.5 if run and push	

Source: own processing

d) IQ Agents Creation Methodology

As previously analysed, each IQ test consists of 8 numbers, except for the last one, with six possible choices. The agent has to analyse these 8 numbers to find the hidden patterns while also reading the 6 possible choices. So their state space is 15 integers. Their possible action is to choose from the 6 available numbers they have as input and they have to find the correct one, while the reward is +1 if they find it correctly and -0.2 if they are wrong, while the IQ test is changed anew when it is solved.

Experimental Results

a) Training Methodology

To begin with, simple agents were developed with specified tasks, such as activating buttons, opening doors, and familiarizing themselves with the communication system. This unique communication system was designed to meet the needs of the OCEAN 5 personality characteristic model. A major element of this system is the 'indicate' action, which allows agents to highlight certain items in the room.

Table 6: Expected agent behaviours based on HiDAC

Personality	Expected Behaviour
Openness	As openness increases, individuals tend to explore more places, ultimately leading them to exit the building at a later time
Conscientiousness and agreeableness	The shortest time occurs when conscientiousness and agreeableness are highest, as agreeable and conscientious individuals tend to be more patient, avoid pushing each other, and exhibit predictable behaviour, favouring cooperation. Conversely, the longest time is observed when both values are minimal.
Extroverts and introverts	Extroverts exhibit quicker movement towards the attraction point, often reaching it in less time. Furthermore, when encountering obstacles such as other agents blocking their path, they tend to resort to pushing them aside in order to achieve their objective.
Neuroticism and non-conscientiousness on panic behaviour	Agents characterized by neuroticism and lower conscientiousness levels demonstrate a tendency to panic more frequently. This behaviour manifests in their inclination to push other agents aside, forcefully navigating through the crowd in a rush to reach the door.

Source: own processing

Initially, agents are entrusted with opening doors by identifying buttons in small, confined areas, then graduating to more complex situations as their expertise grows. Agents are then rewarded for helping other players by recognizing buttons, followed by facing barriers meant to help them learn about queuing and/or to use the ‘push’ action.

After this initial phase, agents undergo independent training on each personality trait, leveraging the established reward model (refer to Table 3). For effective assessment of the agents’ performance, certain behavioural ground truths must be established. As an initial benchmark, experimental results from Durupinar et al. (2011) are utilized for comparative analysis, as delineated in Table 6. Last but not least, the IQ agent was trained to solve the designated IQ tests. To introduce variability in experimentation, the Unity ML-Agents package was used which includes the PPO algorithm and also incorporated the A2C algorithm from the OpenAI platform in the Gym environment.

b) Training Results

After configuring the agents’ reward models, each agent was trained within the same ER environment for a total of 25 million steps, with the results depicted in the following diagrams. Along the horizontal axis, the steps of the training regimen are delineated, while the vertical axis showcases the rewards of the behaviour metrics obtained. To facilitate comprehensive training, the values of the Gaussian distribution Ψ are varied on each occasion, allowing us to train agents across different levels of each trait. Specifically, agents were trained with distributions set to one for positive traits and minus one for negative traits.

Chart 1 illustrates the cumulative rewards garnered by all agents throughout the training process. Notably, the Extrovert emerges as the best-performing agent with a positive aspect, achieving the highest reward of approximately 12,000. Conversely, the non-agreeable agent achieves the highest reward, reaching around 8,000. It is apparent that all agents are capable of navigating and completing the room, albeit with varying reward trajectories. Given the diverse approaches employed by each agent to accrue rewards, direct comparisons between their performances are not feasible.

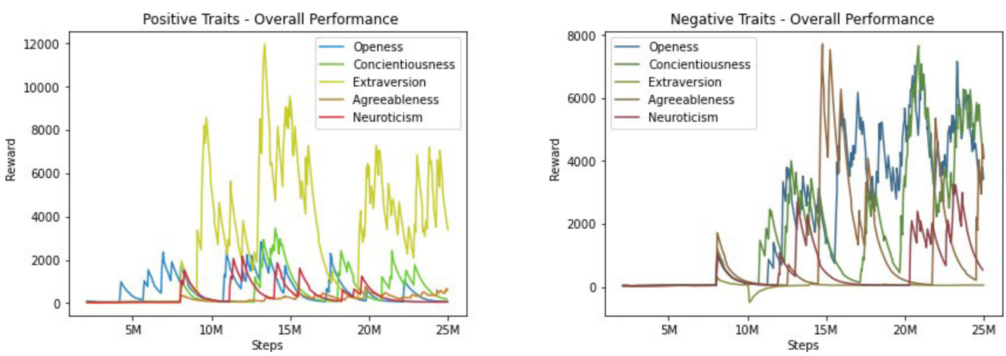


Chart 1: Cumulative rewards for agents

Source: own processing

Chart 2 illustrates the outcomes of the Exploration behaviour exhibited by the Openness agent. Notably, the agent with a positive trait demonstrates consistently high and stable rewards throughout the training process. Conversely, the agent characterized by a negative aspect displays more erratic and negative progress. This stark contrast vividly illustrates the divergent ways in which two agents can interact with the environment, highlighting the profound impact of personality traits on their behaviours and performance.

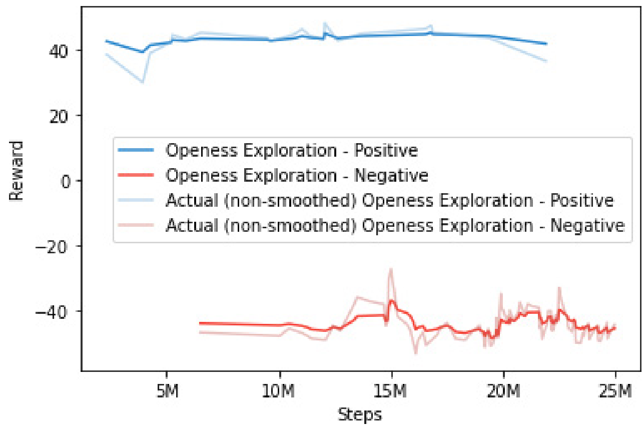
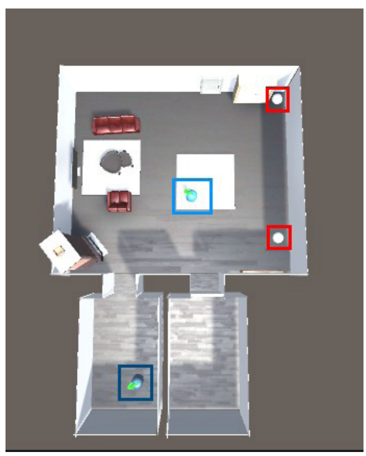


Chart 2: Exploration results for Openness agent

Source: own processing

Picture 4 provides a snapshot of the agents' team environment, depicting two agents both emulating Openness behaviour, albeit with one possessing a positive aspect and the other a negative one. Notably, the agents have successfully unlocked the corresponding doors. However, a clear disparity in behaviour is evident: the agent with a positive Openness aspect is actively exploring the main room, while its counterpart with a negative aspect appears to be proceeding cautiously, displaying hesitancy in venturing out. This visual representation underscores how individual personality traits can influence agents' actions and decision-making processes within the shared environment.



Picture 4: Agents' team environment with positive (light blue) and negative (dark blue) Openness agents and the targets (red)

Source: own processing

In Chart 3 the panic behaviour exhibited by the trained agents possessing both positive and negative aspects are shown. It is evident that the conscientious agent has been significantly impacted, experiencing the most pronounced manifestations of panic behaviour, followed closely by the neurotic agent. Conversely, on the negative spectrum, the non-agreeable agent displays panic behaviour akin to that observed in the introverted and non-open agents. This correlation underscores the destabilizing effect of panic behaviour on efficiency within the room, highlighting its detrimental impact on agent performance.

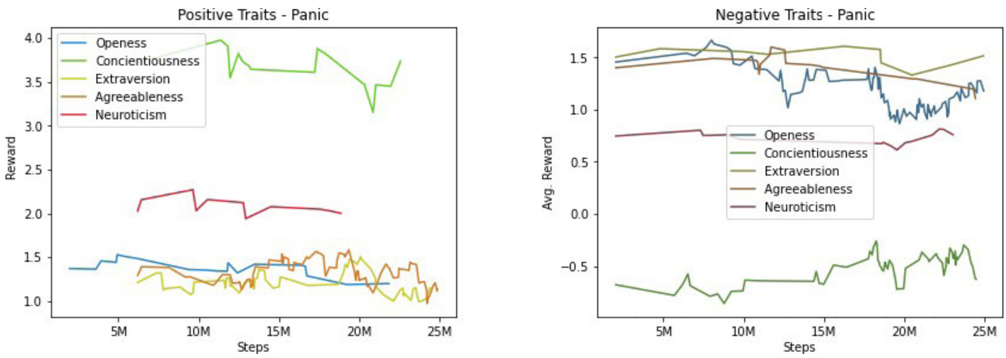


Chart 3: Panic results for all agents

Source: own processing

Chart 4 presents the metrics depicting the impatience levels of the agents. Notably, the extroverted agent exhibits the highest degree of impatience among those with positive traits, while the non-conscientious agent displays even greater impatience with elevated values. This suggests a propensity for these agents to push others when faced with obstacles blocking their path.

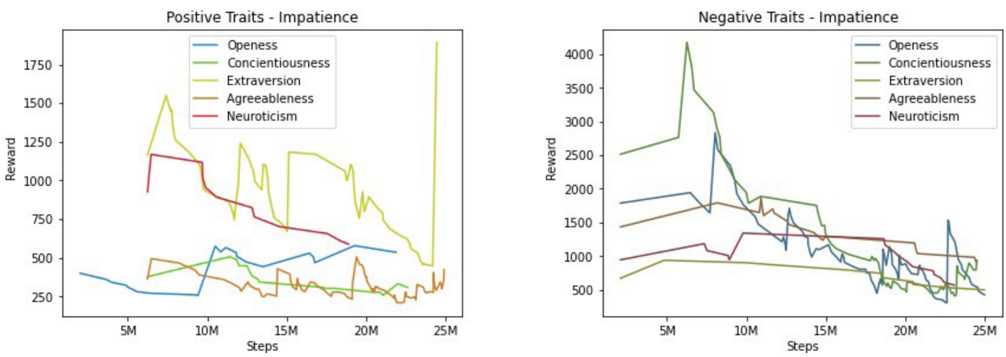
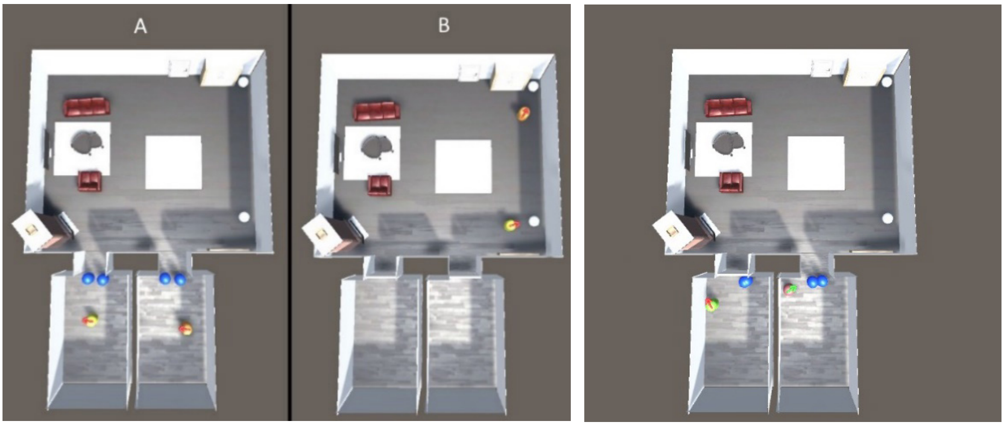


Chart 4: Impatience results for all agents

Source: own processing

In Picture 5, a scenario within a agents' team system featuring two distinct agents is featured. The extroverted agent (depicted in orange) initiates from the right room, whereas the agreeable agent (highlighted in yellow) occupies the left. According to Table 6, these two agents are anticipated to demonstrate high efficiency. True to expectation, they exhibit exemplary collaboration: both agents promptly proceed to locate the button (state A) without resorting to pushing NPCs at the outset. Subsequently, in the main room, each agent efficiently progresses toward the two goals (state B). This synchronized behaviour underscores their adeptness and effectiveness in navigating the environment.

Additionally, Picture 5 depicts a collaborative effort between a conscientious agent (depicted in green) and a neurotic agent (illustrated in red) in their escape endeavour. Remarkably, the conscientious agent is observed opening the door for the neurotic counterpart. However, contrary to expectations, the neurotic agent not only neglects to press the button but also resorts to pushing NPCs within the starting room in an apparent rush to depart. This scenario highlights the contrasting behaviours and priorities of agents characterized by conscientiousness and neuroticism, showcasing the complexities of collaboration within the agents' team environment.



Picture 5: Agents' team environment with Extrovert (orange) and Agreeable (yellow) agents at two different times inside the room (A – close to start; B – close to end) and Conscientious (green) and Neurotic (red) agents at the start of the room
Source: own processing

In Chart 5, the training results of the DRL agent utilizing both the Unity package, and the Gym Environment is showcased. Initially, the agent struggles to find the correct answer, typically requiring several attempts before success. However, as the training progresses, the agent gradually learns to solve the puzzles with fewer attempts, demonstrating consistent improvement in performance over time. Notably, it takes approximately 850 thousand steps for the DRL agent to reliably find the correct answer on the fourth attempt.

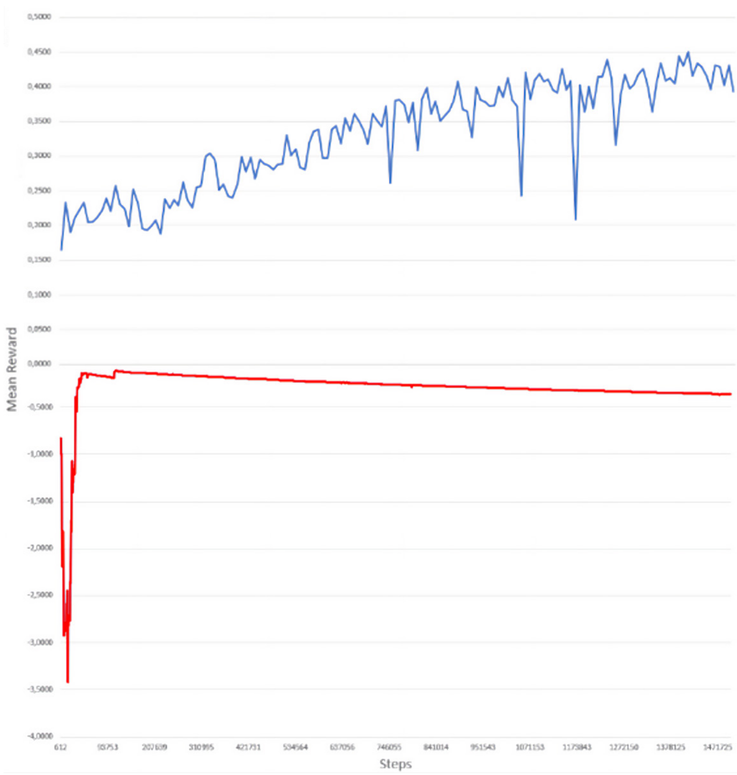


Chart 5: IQ Agents training results – Unity ML-Agents (blue) and Gym environment (red)
Source: own processing

Conversely, the agent trained in the Gym environment initially makes numerous mistakes but swiftly learns to find the correct answer on the second or third attempt, typically within less than 100 thousand steps. However, as training advances, the agent's efficiency plateaus, resulting in less stable rewards.

Overall, these results illustrate the learning capabilities of DRL agents in solving puzzles, showcasing their ability to improve performance over time through iterative training processes. By adapting these elements to different action spaces of other game types, developers can create agents with distinct personality types, tailored to the specific dynamics of each game. The results clearly demonstrate that this methodology successfully produces agents with defined personalities, exhibiting unique and diverse behaviours.

Discussion

The outcomes of the training process validate the initial hypothesis, affirming the agents' capability to emulate human behaviours effectively. This confirmation stems from two distinct perspectives: the rewards obtained and their corresponding values, and the visual inspection of agents' behaviours within the game environment.

Each agent exhibited a diverse array of behaviours, shaped by their respective reward functions, which mimic, in a simplified manner, individual human behaviours. While some agents showcased more intricate actions and gameplay styles due to the coexistence of multiple personality traits, each trait was simplified to its essence. These agents could adapt their behaviours based on varying levels of each trait, reflecting the dynamic nature of human personality. Although personality is inherently complex, each trait and its corresponding behaviour was successfully implemented, culminating in agents that effectively simulate fundamental personality traits.

Furthermore, the agents demonstrated the ability to comprehend mathematical patterns akin to those found in Raven-inspired IQ tests. This ability bodes well for their potential to learn and solve other Raven-like IQ tests, showcasing their aptitude for logical reasoning and pattern recognition involving shapes and colours. These promising results underscore the agents' capacity for learning and adaptation, marking significant progress in the field of artificial intelligence and behavioural simulation.

These findings indicate that these agents could be effectively employed as NPCS in a gaming environment, such as in an ER environment. By integrating these agents into such settings, they can be programmed to exhibit a diverse range of behaviours, characteristics, and decision-making processes, simulating real-world complexities. This approach not only enhances the realism and depth of the gameplay experience but also generates valuable data regarding NPC interactions, responses, and performance.

This data can be systematically analysed to build standard gaming profiles, which define typical patterns of behaviour, decision-making, and outcomes within the game. In a serious game environment – where the goal extends beyond entertainment to include training, education, or skill assessment – these profiles can serve as benchmarks. The generated NPC data can be compared against real player data, providing a meaningful way to assess a player's actions, decisions, and overall performance. Such comparisons could be used to evaluate a player's ability to handle various scenarios, identify gaps in knowledge, and track improvements over time, making these agents a powerful tool in both game design and educational assessment as well a set high scores based on the best performance the NPC can set (e.g. based on the rewards showcased in the previous section).

Lastly, these NPC agents can be adapted for use in a wide range of game types beyond serious games or simulations. They could serve as dynamic characters in role-playing games, strategy games, or even open-world adventure games, where their complex behaviours would enrich the gaming environment by providing more lifelike interactions. Whether it's guiding players through a storyline, challenging them with strategic decisions, or simulating realistic environments, these agents have the potential to enhance immersion and engagement across various genres.

Conclusion and Future Work

This paper introduces intelligent NPC agents that play *MindEscape*, a 3D ER game, and simulate characteristics of OCEAN 5 Personality Traits models. Our game design enables the agents to showcase different playstyles and generate data and standard profiles, regarding their interactions within the room and their behavioural tendencies, alongside their approach to solving Raven-like IQ tests.

Utilizing DRL agents, extensive gameplay data was generated and diverse profiles by emulating characteristics and behaviours associated with personality traits. This approach facilitated the collection of ample and varied data to comprehend the spectrum of human play styles contingent upon the personality model. The analysis of results indicates the agents' capability to emulate these behaviours effectively, encompassing tasks such as navigating complex 3D environments, identifying numeric patterns, and collaborating within a agents' team system.

The successful training of ten agents, with five representing the positive aspect and five representing the negative aspect of each personality trait, revealed substantial diversity in reward values and distinct play styles. These agents displayed a wide range of behaviours, each shaped by their specific personality traits, resulting in varied decision-making approaches. Many agents demonstrated unique, adaptive behaviours, showcasing their ability to collaborate effectively with others in solving room puzzles and facilitating successful escapes. This diversity in behaviour confirms that agents can emulate actions and decision-making patterns based on the OCEAN 5 Personality Traits (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism).

This adaptability suggests that personality-based NPC design could significantly enhance the realism and complexity of gameplay in various settings, particularly in ER environment. Whether acting independently or in a team, the personality-driven actions and problem-solving techniques of these agents create a dynamic, engaging environment that mimics real-world social interactions and decision-making, paving the way for more sophisticated AI-driven gaming experiences and NPCs.

Looking ahead, our implementation will expand to include more complex agents in increasingly diverse environments and game types, featuring new types of tests and puzzles for escape, including various IQ tests and mathematical challenges. Moreover, the development of agents with differing traits will aim to simulate more nuanced and human-like gameplay styles. These future endeavours will further enhance the sophistication and applicability of our approach to creating complex NPC agents and offer a new experience in different kinds of games to the players.

Beyond serious games, these personality-driven NPCs have the potential to greatly enhance other game genres as well. In role-playing games, they can introduce more nuanced character development and interactions, reacting differently to player actions based on their personality. In strategy games, they could take on the roles of teammates or opponents with distinct approaches to problem-solving, resource management, or combat. In open-world or adventure games, NPCs could create a more dynamic and engaging world, with each agent responding in unpredictable ways based on their traits, leading to emergent gameplay.

The ability of these agents to adapt and display varied gameplay styles opens new possibilities for enhancing game realism, player engagement, and challenge across multiple game genres, making them a valuable tool for both entertainment and educational purposes.

BIBLIOGRAPHY

Abt, C. C. (1987). *Serious Games*. University Press of America.

Clarke, S., Peel, D., Arnab, S., Morini, L., Keegan, H., & Wood, O. (2017). EscapED: A framework for creating educational escape rooms and interactive games to for higher/further education. *International Journal of Serious Games*, *4*(3), 73-86. https://doi.org/10.17083/ijsg.v4i3.180

Costa, P. T., Jr., & McCrae, R. R. (2008). The revised NEO personality inventory (NEO-PI-R). In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment: Vol. 2. – Personality measurement and testing* (pp. 179-198). Sage Publications. https://doi.org/10.4135/9781849200479.n9

Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). Gamification in education: A systematic mapping study. *Journal of Educational Technology & Society*, *18*(3), 75-88.

Durupinar, F., Pelechano, N., Allbeck, J., Gudukbay, U., & Badler, N. (2011). How the ocean personality model affects the perception of crowds. *IEEE Computer Graphics and Applications*, *31*(3), 22-31. https://doi.org/10.1109/MCG.2009.105

Ethics guidelines for trustworthy AI. (2019, April 8). Shaping Europe's Digital Future. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Fotaris, P., & Mastoras, T. (2019). Escape rooms for learning: A systematic review. In L. Elbaek, G. Majgaard, A. Valente, & S. Khalid (Eds.), *13th International Conference on Game Based Learning* (pp. 235-243). Academic Conferences and Publishing International Limited.

Goldman, M. (2020, September). *Gamification success stories (Part III): Overview of Michael Sutton's work in game-based learning* [Interview]. tenfold. https://www.researchgate.net/publication/346449392_Gamification_Success_Stories_Part_III_Overview_of_Michael_Sutton's_work_in_Game-Based_Learning

Jang, K. L., Livesley, W. J., & Vernon, P. A. (1996). Heritability of the big five personality dimensions and their facets: A twin study. *Journal of Personality*, *64*(3), 577-592. https://doi.org/10.1111/j.1467-6494.1996.tb00522.x

Liapis, G., Lazaridis A., & Vlahavas, I. (2021). Escape room experience for team building through gamification using deep reinforcement learning. In P. Fotaris (Ed.), *15th European Conference on Game Based Learning* (pp. 472-479). Academic Conferences.

Liapis, G., Zacharia, K., Rrasa, K., Liapi, A., & Vlahavas, I. (2022). Modelling core personality traits behaviours in a gamified escape room environment. In C. Costa, F. Luz, & N. Fachada (Eds.), *16th European Conference on Games Based Learning* (pp. 723-731). Academic Conferences International Limited. https://doi.org/10.34190/ecgbl.16.1.602

Raven, J. (2000). The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology*, *41*(1), 1-48. https://doi.org/10.1006/cogp.1999.0735

Robson, K., Plangger, K., Kietzmann, J. H., McCarthy, I., & Pitt, L. (2015). Is it all a game? Understanding the principles of gamification. *Business Horizons*, *58*(4), 411-420. https://doi.org/10.1016/j.bushor.2015.03.006

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). *Proximal policy optimization algorithms*. arXiv. https://doi.org/10.48550/arxiv.1707.06347

Tsaousis, I. (1998). The traits personality questionnaire (TPQue): A Greek measure for the five factor model. *Personality and Individual Differences*, *26*(2), 271-283. https://doi.org/10.1016/S0191-8869(98)00131-7

Uzieblo, K., Verschuere, B., Van den Bussche, E., & Crombez, G. (2010). The validity of the psychopathic personality inventory – Revised in a community sample. *Assessment*, *17*(3), 334-346. https://doi.org/10.1177/1073191109356544

Vergne, M. J., Smith, J. D., & Bowen, R. S. (2020). Escape the (remote) classroom: An online escape room for remote learning. *Journal of Chemical Education*, *97*(9), 2845-2848. https://doi.org/10.1021/acs.jchemed.0c00449

Appendix A: Personality Traits Behaviours Equation Based on Durupinar et al. (2011)

Openness
Trained
$\beta_{tr} = 1 \text{ if } \Psi^N \geq 0.5 \text{ else } 0$
Exploring
$\beta_{ex} = 10 * \Psi^O$
Conscientiousness
Panic
$\beta_{pa} = W^{CP} * f(\Psi^C)$
where $f(\Psi C) = -2 * \Psi^C + 2 \text{ if } \Psi^C \geq 0 \text{ else } 0$
Impatience
$\beta_{imp} = W^{CI} * (1 - \Psi^C)\ 0$
Right Preference
$\beta_{rp} = 1 \text{ if } P(right) \geq 0 \text{ else } 0$
Where $P(right) = 0.5 \text{ if } \Psi^C < 0 \text{ else } W^{CR} * \Psi^C$
While $W^{CP} + W^{CI} + W^{CR} = 1$
Agreeableness
Impatience
$\beta_{imp} = W^{AI} * (1- \Psi^A)$
Pushing
$\beta_{pu} = 1 \text{ if } W^{AP} * (1- \Psi^A) \geq 0.5 \text{ else } 0$
Right Preference
$\beta_{rp} = 1 \text{ if } P(right) \geq 0 \text{ else } 0$
where $P(right) = 0.5 \text{ if } \Psi^C < 0 \text{ else } W^{AR} * \Psi^C$
Wait radius
$\beta_{wr} = 0.25 \text{ if } \Psi^A \in [0,1/3) \text{ else } 0.45 \text{ if } \Psi^A \in [1/3, 2/3] \text{ else } 0.65 \text{ if } \Psi^A \in (2/3,0]$
Wait timer
$\beta_{wt} = 1 \text{ if } \Psi^A \in [0,1/3) \text{ else } 5 \text{ if } \Psi^A \in [1/3, 2/3] \text{ else } 50 \text{ if } \Psi^A \in (2/3,0]$
While $W^{AI} + W^{AP} + W^{AR} = 1$
Extroversion
Leadership
$\beta_{le} = W^{EL} * \Psi^E$
Communication
$\beta_{co} = 1 \text{ if } \Psi^E \geq 0.5 \text{ else } 0$
Impatience
$\beta_{im} = W^{EI} * f(\Psi^E)$
where $f(\Psi^E) = 2 * \Psi^E - 1 \text{ if } \Psi^E \geq 0 \text{ else } 0$
Pushing
$\beta_{pu} = 1 \text{ if } W^{EP} * \Psi^E \geq 0.5 \text{ else } 0$
Walk speed
$\beta_{wp} = \Psi^E + 1$
Gesture
$\beta_{ge} = 10 * \Psi^E$
Personal space
for Agent i and j on a queue
$\beta_{ps} = 0.8 * f(i,j) \text{ if } \Psi^E \in [0,1/3) \text{ else } 0.7 * f(i,j) \text{ if } \Psi^E \in [1/3, 2/3] \text{ else } 0.8 * f(i,j) \text{ if } \Psi^E \in (2/3,0]$
Where $f(i,j) = 1 \text{ if } i \text{ before } j \text{ else } 0.4/0.7$
While $W^{EL} + W^{EI} + W^{EP} = 1$

Neuroticism

Leadership

$\beta_{le} = W^{N\,L} * (1 - \Psi^{N})$

Panic

$\beta_{pa} = W^{N\,P} * \Psi^{E}$

While $W^{N\,L} + W^{N\,P} = 1$