More than buttons on controllers: engaging social interactions in narrative VR games through social attitudes detection

Georgiana Cristina Dobre
c.dobre@gold.ac.uk
Goldsmiths University of London
London, UK

Marco Gillies
m.gillies@gold.ac.uk
Goldsmiths University of London
London, UK

David C. Ranyard
dave@dreamrealityinteractive.com
Dream Reality Interactive
London, UK

Russell Harding
russell@maze-theory.com
Maze Theory
London, UK

Xueni Pan
x.pan@gold.ac.uk
Goldsmiths University of London
London, UK

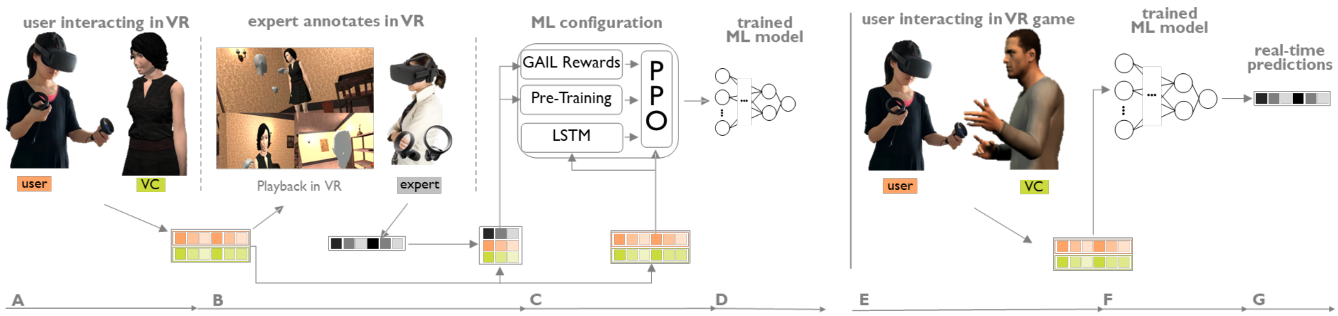

Figure 1: Pipeline for detecting human-defined social attitudes, including immersive data collection (user interaction (A) and expert annotating (B)) for training the machine learning model. This takes place by pre-training the model, creating Generative Adversarial Imitation Learning (GAIL) rewards for the reinforcement learning algorithm Proximal Policy Optimisation (PPO) that also uses a temporal memory called Long Short-Term Memory (LSTM) algorithm (C). This process exports a trained ML model (D). In a user-VC interaction (E), the trained model (F) detects in real time the human-defined social attitude (G) which could be used in different scenarios.

ABSTRACT

People can understand how human interaction unfolds and can pinpoint social attitudes such as showing interest or social engagement with a conversational partner. However, summarising this with a set of rules is difficult, as our judgement is sometimes subtle and subconscious. Hence, it is challenging to program agents or non-player characters (NPCs) to react towards social signals appropriately, which is important for immersive narrative games in Virtual Reality (VR). We present a collaborative work between two game studios (Maze Theory and Dream Reality Interactive) and academia to develop an immersive machine learning (ML) pipeline for detecting social engagement. Here we introduce the motivation and the methodology of the immersive ML pipeline, then we cover the motivation for the industry-academia collaboration, how it progressed, the implications of joined work on the industry and reflective insights on the collaboration. Overall, we highlight the industry-academia collaborative work on an immersive ML pipeline for detecting social engagement. We demonstrate how creatives could use ML and VR to expand their ability to design more engaging commercial games.

IVA '22, September 6–9, 2022, Faro, Portugal

ACM ISBN 978-1-4503-9248-8/22/09…$15.00
https://doi.org/10.1145/3514197.3551496

CCS CONCEPTS

• **Software and its engineering** → **Interactive games**; • **Human-centered computing** → **HCI design and evaluation methods**; • **Computing methodologies** → *Intelligent agents.*

KEYWORDS

Artificial Intelligence, Expressive Body Language, Gaming, Human-Computer Interaction, Virtual Agents, Virtual Reality

ACM Reference Format:
Georgiana Cristina Dobre, Marco Gillies, David C. Ranyard, Russell Harding, and Xueni Pan. 2022. More than buttons on controllers: engaging social interactions in narrative VR games through social attitudes detection. In *ACM International Conference on Intelligent Virtual Agents (IVA '22), September 6–9, 2022, Faro, Portugal.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3514197.3551496

1 INTRODUCTION

Complex human behaviours exhibited in everyday social interaction are hard to recognise automatically and therefore to use as mechanics in videogames. As a result, players often find themselves driving a social interaction in a videogame by choosing what to do from a menu (see Section 4.1 for examples). In Virtual Reality (VR), this could break the plausibility illusion [Slater 2009] and lead to break-in-presence [Slater and Steed 2000], which takes the players back to the real-world and significantly reduces the level of immersion.

In this work, we show a collaboration between two game studios and academia. It explores a novel pipeline in game design, combining Machine Learning (ML) and VR, with the aim to make social interactions in VR narrative games more engaging, immersive, and inclusive (in the sense that it will appeal to a broader audience than current video games).

One of the most promising uses of body movement in VR is the social interaction with Virtual Characters (VCs), or Non-Player Characters (NPCs). In face-to-face interactions with people, we use our bodies extensively as non-verbal communication (colloquially called 'Body Language'), including actions such as gaze (eye contact), gestures, posture and the use of personal space. VR opens the possibility to use these social cues as first-class elements of gameplay and thus creating much richer social experiences.

There are clear benefits to replace traditional *explicit interactions* (selecting an option by clicking a button) with *implicit interactions* (social attitudes expressed via body language), where the player's non-deliberate, *implicit* actions are inputs for a game [Schmidt 2000]. This is in particular important for maintaining the plausibility illusion in player-driven narrative games in VR where players could engage with NPCs naturally.

Here, we propose a pipeline for detecting social attitudes with a use-case on social engagement. This can be used as implicit interactions in a narrative VR game. The detected social engagement can then be used to trigger different behaviours in the NPC or in the game environment itself, influencing how the game continues. These triggers are to be decided by the game designers and game creators, based on how they envision the game and the gameplay. For instance, if a player is detected to be socially engaged with an NPC, they could gain higher trust scores from this NPC and could display animations that reflect a higher level of social engagement in return. Hence, we focused on the social engagement detection part, the NPC's response to the user's social engagement in the game being out of our scope.

In this work, we treat the definition of social attitudes as a creative process driven by the judgement of a game designer. Thus, social attitudes are concepts without an explicit definition or ground truth and defined solely through expert creative judgement. This type of interaction design concept will be increasingly common as VR becomes a medium used by creative practitioners and which attempts to use more complex and subtle aspects of human behaviour.

2 MOTIVATION

Maze Theory and Dream Reality Interactive are two games companies focused on developing engaging immersive games in VR and AR. One of Maze Theory's new development is a narrative VR game on the popular British TV-show Peaky Blinders (www.peakyblinders.tv). The Peaky Blinders show is a crime drama series showcasing strong and complex characters.

Maze Theory plans to rebuild the characters in the game by developing them to be responsive, engaging and believable. The current technology of creating the interaction for NPCs in games relies on pre-determined scripted sequential events, which is limited given the kind of NPCs in Peaky Blinders VR game. Considering the challenges of creating NPCs for VR games, there's need for expertise in this area. More complex methods, such as Machine Learning-based, are primarily used in games for graphics generation or navigation rather than emotion recognition or nuanced one-to-one VR interactions. Research from academia can be used to create more engaging experiences and enhance the user's feeling of presence and immersion. Thus, the motivation for this collaboration lays into a deeper understanding of the human interaction that can be applied to NPCs in VR games.

The player's sense of presence is decisive in VR games; thus this collaboration focuses on the insights of naturally reactive NPCs, what is expected from an NPC in the presence of a player, how to translate the NPC's awareness of the player in the game, and finally, how are all of these integrated with the entertaining aspect of a game. The overall aim is to get insights into more believable NPCs/VCs in VR, maintaining the player's presence as if they are interacting with someone in real life. This collaboration initiated the vision of a game narrative driven by the player's interaction with the NPC. Hence we consider the emotion, body language, and nonverbal behaviours based on the actions in the scene.

Our approach focused on designing and developing a pipeline for detecting social attitudes in NPC-player social interactions. We decided that the most suitable social attitude is social engagement. We came to this conclusion because the user's interactions with the NPCs in Peaky Blinders game will depend on the user's overall actions. They will influence the game progression and the relationships in the game.

To strengthen the industry-academia collaboration, the lead author in this paper worked on this project full time in one of the collaborating game studios (Dream Reality Interaction). This enhanced the collaboration during and after the project. It led to a better blend of academic practises in the games industry, and at the same time, it exposed the researcher to the fast paced environment of game design and development. The full-time work took place at the beginning of the project, covering the pipeline and the study design process, the development of the pipeline, the data collection and the initial part of the ML training. The work from the game studio, allowed for the pipeline's design and development to happen from within the games industry. This let to a strong collaboration on the pipeline development ensuring that the work fulfils the game studio's needs.

3 CONTRIBUTIONS

The academia-industry collaboration led to a creative director-focused ML pipeline for social engagement detection. This pipeline provides three principle novel contributions. First, we designed and conducted an experimental study of an immersive data collection process in which participants listened to an NPC's monologue (prepared by professional writers) in a VR environment closely resembling a real game social interaction.

Secondly, we developed an immersive environment where game designers could annotate the captured data, identifying instances of the social engagement. This VR environment placed the annotator in the same virtual space as the participant and the NPC, enabling them to watch the interaction as if it was a real-life conversation. This allowed them to make the most effective use of their social cognition and this also created an artist-friendly environment for data annotation, which was close to real gameplay experiences. The latter turned data annotation from a technical task to one that benefits from an interaction design skill.

Finally, with our pipeline, we were able to train an ML model to detect implicit social engagement in VR interactions with 83% accuracy. Specifically, we used a reinforcement learning algorithm with imitation learning rewards from examples set by human experts.

4 RELATED WORK

4.1 NPCs in narrative games

Narrative or story-driven games are those with a clear storyline where the players' actions are based on the story and can influence it [Ip 2011]. In these type of games, the game mechanics are not only centred around actions performed in the game, but also in the story behind the gameplay. The narrative function in a game creates compelling and engaging play as it borrows aspects from other forms of narrative media (such as film and literature), adding emotional depth to the player's experience.

In many narrative games, the NPCs (or VCs) are a core element. Often players can interact with them (fight/get help from) or even have a dialogue with them as part of the gameplay. The outcome of the interaction often leads directly on to the next actions available, making the interaction itself part of the game mechanics. Hence, the narrative games are designed around the overall story and the player's interaction with the NPCs. Examples of games that fall into this category include: *The Walking Dead Series* (www.skybound.com/telltales-the-walking-dead-the-definitive-series), *Heavy Rain* (www.quanticdream.com/en/heavy-rain), *Mass Effect* (www.ea.com/en-gb/games/mass-effect) or *L.A. Noire* (www.rockstargames.com/lanoire). In most cases, these games are non-VR and dialogues with the NPCs involve players selecting phrases from a pre-defined list, using the buttons from mouse, keyboard, or joysticks.

Recent years have seen the rise of VR games which push the game engagement to the next level. Some of these games applied the classic game mechanics and interaction methods directly from non-VR games onto the VR ones (*Hellblade: Senua's Sacrifice VR* www.hellblade.com), others attempted to adapt some of the interactions to 3D controllers. For instance, in *Moss* (www.playstation.com/en-gb/games/moss-ps4), players could use the PlayStation controllers to navigate and interact, and are directly involved in the narrative by controlling the main character (a young mouse called *Quill*) from a third-person perspective. Although these games are more immersive than non-VR games, using VR does not necessarily lead to improved user experience (e.g., simulation sickness in [Christensen et al. 2018]).

We argue that it is key to enable natural interaction utilising the richer inputs VR offers. Because users can move freely in VR, the interaction in these games does not have to be restricted by the game controllers. The user's large and diverse range of inputs can be manipulated to design interactions with NPCs that are closer to the ones taking place in real life. This aspect helps maintaining the user's plausibility illusion, which means that the user's experience of interacting with an NPC (VC) is similar to an interaction that happens face to face with a real person.

There are popular VR games that make good use of natural interactions, such as *Beat Saber* (www.beatsaber.com) or *SuperHot VR* (www.superhotgame.com/vr). However, most of them are not centred on a story, nor the interaction with NPCs.

It is difficult to develop narrative games in VR with natural social interactions. This is because the natural interactions with NPCs are more complex than the interactions in non-narrative settings (such as slicing cubes with lightsabers- in Beat Saber). Other VR games, such as *Half-Life Alyx* (www.half-life.com/en/alyx), implement ways of interacting with the environment that are very close to how people do in daily life. Being able to open doors by pushing them, manually reloading weapons, crawl and freely move around, enhances users' feeling of presence. However, most of the games like this one, rely on core mechanics such as shooting or fighting, making them violent. Having these violent behaviours happen in VR can have a strong and profound effect on the players' emotion and behaviour ([Bailenson 2018; Wilson and McGill 2018]), thus excluding users less interested in violent or action-based games.

The games industry are trying to find other ways of designing interactions and other game mechanics that would better fit the VR medium. The industry-academia collaboration aims to aid the creation of first-person VR narrative games that make use of the VR technology and that is not centred around traditional game actions but rather on social interactions with an NPC/agent.

4.2 Detecting social engagement

Being able to detect social attitudes in narrative games during the player-NPC interactions can lead to implicit interaction and new game mechanics. The industry-academia collaboration aims at building a tool to detect social engagement. Here, it broadly refers to the social engagement one shows in social interactions linking it to the action of showing interest, supporting and paying attention.

Hence, social engagement is a complex and subjective social attitude that is difficult to be described using concrete rules. Humans, on the other hand, have the ability to easily identify and interpret complex non-verbal behaviours, even from a still image [Vinciarelli et al. 2011]. Since this understanding is implicit, and we are designing a machine learning process based on creative judgements not on an objective definition, we do not formally define social engagement (there are also many engagement definitions [Glas and Pelachaud 2015]). Instead, the concept emerges implicitly from the annotator's judgement of participants' behaviour.

Social engagement is an important aspect to consider during user-VC interactions. As with other social attitudes, the VC should adapt its behaviour with a change in the engagement level. This has been researched on many occasions, for instance in [Bohus and Horvitz 2014; Dhamija and Boult 2017; Gordon et al. 2016;

Figure 2: Example of user interactions with the VC (left) and its correspondent expert's annotation view (right). On the left figures: a user pats on the VC's shoulder (top) and a user looks at a piece of furniture (bottom); on the right: the expert annotator's view of the interactions from the left, along with extra information of the user's and the VC's camera views at the bottom of each annotator's view. These extra views can be activated or deactivated by the annotator.

Woolf et al. 2009]. They propose methods that tackle engagement in interactions, however, they disregard the user-VC interaction dynamics loop. [Dermouche and Pelachaud 2019] include this loop in their work, detecting the engagement from dialogue videos on a 5-level engagement scale.

To detect social engagement in social interactions, we propose a pipeline based on *imitation learning*. We introduce a method to integrate natural social interaction aspects as game mechanics in narrative VR games. The method is based on synchronised data from both interaction participants (VC and the user), as it would happen in the final game. Through this, the game can detect social engagement and trigger an action that would make the game progress without the user's explicit input.

5 THE PROCESS

In this Section, we describe how to detect social engagement between a user and a VC in an immersive VR scenario using the ML pipeline from Figure 1. During the interaction, the VC carried out a monologue about her family and her life. The monologue was written during a writers room organised by a national centre for immersive storytelling (see Table 1 and the supplementary video in [Dobre et al. 2022]).

The study took place in VR and contained three stages based on the user instructions, which aimed to trigger both high and low social engagement behaviours in users. All tasks were designed from industry-academia brainstorming sessions to represent typical gameplay (Figure 4). In the first VR stage ($S1$) the user was told to interact with the environment and the VC as they would do in a gameplay, allowing us to study the range of different behaviours that participants would perform without a prompt. In the second VR stage ($S2$), the user received instructions to gain the VC's trust, representing the kind of task players would be given in the game. This VR stage aims to record mostly high social engagement data. For the final VR stage ($S3$), the user was asked to explore the room, players being familiar with this task from other games. The interaction from $S3$ aimed to produce primarily low social engagement data. For an example of the participants' behaviour in each part, see videos in supplementary material from [Dobre et al. 2022].

5.1 User data collection in VR

There were in total 13 participants, 9 males and 4 females, aged between 20 and 46 years and an average of 32 years old. All participants voluntarily agreed to take part in the experiment and signed a consent form. The whole process received ethical approval.

The data collection took place in two batches because of time and participants availability restrictions. The first batch was with 6 participants and the second with 7. The only difference between the first and second batch is in the VC's location and gaze direction (see Figure 3).

The term *session* refers to each time the participant took part in the virtual scenario (regardless of the VR stage), hence, there are three sessions for each participant. There is a missing session from S3 in the second batch due to a software error, resulting in a total of 38 sessions, with 18 sessions from the first batch (6×3) and 20 from the second ($7 \times 3 - 1$). In total, the time spend in the VR environment by all participants is approximate 190 minutes (38 sessions $\times$ 5 minutes per session).

The experiment ran on Unity3D and we collected data from both users and the VC (see Table 2). In total, over 108000 frames of multi-modal data were used to train and evaluate the ML model.

5.2 Human annotations in VR

An expert human annotator watched a playback of the user interacting with the VC and annotated their interaction (see Figure 2). As social engagement is a very subjective term and it has many definitions (see [Glas and Pelachaud 2015]), a human annotator marked the data without directly defining social engagement. In this case, the annotator implicitly defined social engagement by annotating it during the user-VC interactions.

This makes the definition of social engagement a human and creative judgement. In game companies, this annotator role is taken by their creative director, making the labelling itself part of the creative process of interaction design, analagous to the development of game mechanics. A game mechanic and the machine learning system both identify player behaviours that should have certain impacts on the game, and both, therefore, should be designed and tweaked by game designers.

The annotation occurs in VR, in the same room environment that was used for the data collection (see Figure 2). The annotator can see a representation of the player with the VC and is able to make judgements based on the interaction between them. The player data is shown in the simplified form (a grey head and hands in Figure

Table 1: A snippet of VC's monologue. The text in italic represents the scriptwriter's indications. As an interactive monologue, the user was directly addressed to in sections such as *Do you think they would have found a new home?* For the monologue animation, see supplementary videos in [Dobre et al. 2022]

Wistful monologue spoken with a sombre tone.
`VC: That's the only place we could laugh freely. The park with the rose finches. They've built apartments on it now. No longer can I ever go there. I wonder what happened to all the finches? Maybe they found a new home.`
VC stares directly at the player again, her brow slightly crumpled.
`VC: Do you think they would have found a new home?`
VC shakes her head briefly and her shoulders slump over a little bit.
`VC: No, they're like me, still looking for somewhere else to call home. I often imagine them happy[...]`

Figure 3: Virtual Character's viewpoint location in each data collection batches (batch 1- top, batch 2- bottom)

Table 2: Data was recorded from participants and VC; The head and hands are relative to the corresponding root of each VC and the user; 3D Vectors represents the X, Y and Z components in a vector data structure; The Quaternion represents the X, Y, Z, W rotation components

Information Recorded	Data Type
User's head position	3D Vector
User's head rotation	Quaternion
User's left- and right-hand position	3D Vector
User's left- and right-hand rotation	Quaternion
User's main head root position	3D Vector
User's main head root rotation	Quaternion
User's left & right index and hand triggers	Float
User's headset velocity & angular velocity	3D Vector
VC's head position	3D Vector
VC's head rotation	Quaternion
VC's left- and right-hand position	3D Vector
VC's left- and right-hand rotation	Quaternion
VC's main root (hip) & chest position	3D Vector
VC's main root (hip) & chest rotation	Quaternion

2), which only displays the tracking data that was gathered during the data gathering phase. This simplified format is important, as a more complex representation might cause the annotator to read emotional behaviours into the player that were not present in the data, thus leading to misleading annotations[Gillies et al. 2015].

During the VR interaction playback, they label the behaviours identified in players which are related to the presence or absence of social engagement. Thus when real-life players exhibit those behaviours during gameplay, certain events (VC behaviours, or change of game environment) could be then triggered. The annotator labelled the sessions in random order without knowing which VR stage or user they were annotating.

To assure the annotator had rich social interaction information, they could access the user's and the VC's camera view (showing their current viewpoint). This allowed the human annotator to have access to exactly what they were viewing at any time while being in the same place as the user and the VC. An example of this is seen in Figure 2.

5.3 The ML component

We trained the model using imitation learning with the Unity ML-Agents platform (v0.11) and the reinforcement learning algorithm Proximal Policy Optimization (PPO). PPO provides positive rewards for performing the desired behaviours and negative ones for the non-desired behaviours. In this case, to mimic an imitation learning scenario, the rewards are calculated using Generative Adversarial Imitation Learning (GAIL).

We hypothesize PPO improved performance through pre-training and adding a temporal memory (via a recurrent neural network: Long Short-Term Memory, LSTM). The behaviour learned from pre-training influences the action taken by PPO. At the same time, The temporal memory takes into account past actions, hence the algorithm considers past behaviour and current actions when deciding what to do next (what action to take). We hypothesize this because the behaviour that needs to be learnt is complex and temporal. We compare these models with those without pre-training and LSTM (using random initialisation instead of pre-training and a standard feedforward network instead of LSTM).

There is strong evidence in the literature that certain behaviour aspects (such as body posture or gaze) are linked to social engagement ([Mota and Picard 2003; Sanghvi et al. 2011]). Based on this, we trained the model with psychologically-based features, as shown in Table 3. Since these are calculated from the raw data we collected, we call them derived data. We trained and compared ML models with the derived features or the raw data.

Table 3: Derived input data. These are calculated based on the raw data detailed in Section 5.1 Table 2

Description	Datatype
Distance between the user and VC, based on Hall's personal space [Hall 1966], mapped between [0, 1]	float
User's facing direction: the angle between VC's head rotation and user's head rotations divided by 180	float
Interaction with objects: data from the controllers' trigger (the trigger allows objects interaction)	float
User's headset velocity	$3D$ vector
User's headset angular velocity	$3D$ vector

The user's distance from the VC is calculated based on Hall's personal space [Hall 1966]. See [Dobre et al. 2022] for more details on this feature. Both models (derived and raw data) use the human's annotations as ground truth data, and its output is a discrete binary value. It shows the current user's social engagement at each frame: 1 high social engagement, or -1 low social engagement. These two options aim to mimic the human's ratings of low/high social engagement during the annotation.

We compared two additions to the PPO ML structure: pre-training and a temporal memory via LSTM. We randomised the dataset sessions and divided it into three folds of 13, 13 and 12 sessions each, for a 3-fold cross validation; the training data consists of two folds while the remaining one represents the evaluation data.

Our hypothesis that pre-training and adding a temporal model into the ML configuration is validated. The model configuration based on a reinforcement learning algorithm (PPO) with imitation learning rewards (GAIL) implementing a temporal memory (through LSMT) and a pre-training algorithm showed the highest accuracy (83.4%) and F1-score (84.1%). The other model configurations perform poorly, the outcome being a rapid change between high and low social engagement values in a short period of time. These results are from ML model trained with derived data.

When comparing the models trained with derived data to the ones with raw data we see using derived data yields the best result. The model based on raw data is not able to generalise to different VC positions. This is a problem as it is very common in games to have VCs that would move in the environment. The use of psychologically inspired derived features is therefore a better approach. For a detailed discussion on the ML results, see [Dobre et al. 2022].

6 INDUSTRY OUTCOMES

The most commonly used method to develop NPCs in games is via simple rule-based methods and state machines. This usually applies to NPC for PC or console games seen on 2D monitors. The current development for NPCs in VR games uses a similar approach, which brings some shortcomings. The player in VR games experiences the events happening to them from the first person perspective, embodying the character. In non-VR games, the player usually sees the characters from the third person perspective. Even with first person perspective non-VR games, game creators need to spend many resources to make the player identify with the main character. Developing NPCs for VR games means that the player embodies their character and this brings challenges related to maintaining the player's presence and implementing natural interactions with NPCs.

The academia-industry work was very collaborative. Primarily we focused on how to bring the research knowledge from the academia into the commercial area, to fill the gap and apply the research results to enhance the NPC behaviour in games. We worked closely from a development point of view, tackling the problems of creating NPCs and the game production problems. We met regularly for project meetings and brainstorming sessions, and collectively developed the pipelines for detecting social attitudes for NPCs/VCs (see Fig. 4). The lead author of this work was employed on a full-time internship programme for one of the game studios. This improved the exchange of ideas and allowed for many interactions and insights with different perspectives. It also helped diversifying the office environment and led to a better exposure of the academic practices and processes directly into the industry.

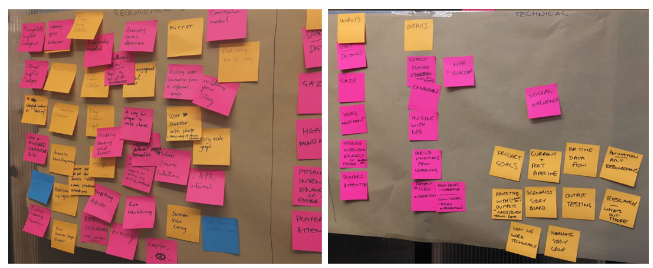

Figure 4: Visuals of notes from brainstorming sessions between the academia and the games industry.

The outcome was slightly different on each game company. During the collaboration we chose the social engagement case study, though the pipeline was also applied for a prototype karaoke-style VR game by Dream Reality Interactive. In this prototype, the user would sing along an NPC singer. The NPC singer will change their behaviour depending on the social attitude of the player in real-time. Likewise, in this prototype, there was no definition for the social attitude. The annotations were based on the way the players sang, moved, performed, and how involved they were in the experience. It was noteworthy that the pipeline designed for an early 20th century narrative game (Peaky Blinders VR) also applied to a modern karaoke game. Unfortunately given the timing and the company priorities, the prototype didn't turn into a full game, however it is projected as future work.

Secondly, although the pipeline yielded good results, Maze Theory ran into difficulties implementing it into the new production game. This was due to performance issues, having to consider both the game technical performance and how reactive the NPC behaved in the game. It is more straightforward to make a background NPC aware of and responsive of the player but it's challenging to blend NPCs based on rich MOCAP data with AI-driven characters. Maze Theory has this challenge on the road map for investigation, aiming to bring the two NPC types together more dynamically, particularly for social VR which is on the future plans. The aim is to engage in industry-academia collaborations, to perform more iterations on this vision, and to learn more on creating AI-driven NPCs for immersive VR games.

The collaboration impacted the industry on a few levels. Firstly, even though the pipeline was not implemented in the Peaky Blinders VR game by Maze Theory, the collaboration and project informed and influenced the decisions taken when designing the NPCs for this game. The collaboration also strengthen the relationship between both game companies and Goldsmiths University. This led to regular student internships in the Dream Reality Interactive game company, helping bring students straight from the academia in the games industry to better bridge the two areas.

7 DISCUSSION AND CONCLUSION

This paper presents collaborative work between two game companies and an university research lab. We developed a pipeline with an immersive data collection and annotation in VR for training an ML model. The pipeline supports the games industry creative design process and to be integrated into production-ready VR games for the consumer market. It is used to train an ML model to detect social engagement using a reinforcement learning (PPO) approach with rewards based on an imitation learning algorithm (GAIL).

The work contributes to the field of socially responsive VCs, offering a design by example tool for immersive ML, to detect social engagement (and possibly abstract social attitudes) in VR social interactions. This could be useful in designing social interactions in VR games or in other immersive experiences (simulations, training, social platforms), where the user can interact with the VC using their own bodies, as they do in everyday life. This opens opportunities for novel input interactions, game mechanics or VC's behavioural models that are related to the rapport/empathy between the user(s) and the VC.

The collaboration bridges the academia and the games industry impacting both games companies on different areas. For Maze Theory, the project informed their decision for the NPCs in their new Peaky Blinders VR game and helped setting the vision for AI-driven characters in VR games over the coming projects. Dream Reality Interaction implemented the proposed pipeline in a different VR karaoke game and are building a strong collaborative link with academia via internship from the university.

The industry-academia collaboration successfully developed an immersive ML pipeline to detect social engagement in an interaction between players and NPCs. The detection builds into more implicit game mechanics and less button-pressing during social interactions with NPCs in VR games. The projects impacted the games industry, influencing both companies production road-maps. The academic group also benefited from the collaboration with the industry, through being exposed to the industry practises and understanding how the research outcomes can be implemented into commercial products.

To conclude, we can offer some insights from this project, and in particular the industry collaboration, which can help the further development of virtual agents both in industry and academia.

The first is that ML can be used effectively in a game development pipeline, but it is important that the process of ML (the gathering and annotation of data) is part of an overall creative process. It is no longer a neutral, objective process of data gathering, that might be used to from a scientific machine learning background, but a design process involving human judgement. This is showed in the data gathering portion, in which participants were given carefully designed instructions that aimed to guide them towards specific types of gameplay behaviour. It was even more important in the annotation, which was part of an interaction design process that used human judgement to define social engagement. It was a game designer's judgement that determined which player's behaviours should elicit certain responses in a virtual character (NPC), just as it would be the case in the design of traditional game mechanics.

A second insight is that non-verbal communication can form the basis of an interesting gameplay mechanic that simulates realistic social interaction. It was decided early in the process that verbal interaction would not be suitable. A traditional game approach to dialog in which players select options from a menu would not feel sufficiently like a real conversation to elicit a strong enough sense of presence. On the other hand, natural language understanding technology was not considered ready for deployment, with cloud technologies being too slow and other technologies likely to perform badly across the diverse voices and accents typical of the global market for modern video games. On the other hand, casting the player in the role of an attentive listener can make them feel like they are in a real conversation and that the character responds to their behaviour. This type of social behaviour is complex, embodied and tacit, which means that machine learning is essential as it would be extremely hard to recognise with hand coded algorithms.

This implies an important shift in game player behaviour. In current games, characters are not aware of players during dialog and players can typically perform other tasks during a monologue by an NPC, for example, exploring the room or even throwing the furniture around, behaviours that would be seen as extremely rude in a real conversational setting. Forcing players to engage with characters appropriately in a socially meaningful interaction is likely to fundamentally change how they view those characters, creating greater social connection with them. Players who outwardly show respects to characters are likely to respect them more inwardly. We hope that this work will help lead to a new generation of games based on strong social connections, and in which the primary mode of interaction is not violent but pro-social.

8 ACKNOWLEDGMENTS

The project was supported by *Innovate UK* grant $TS/S02221X/1$ and partly supported by grant $EP/L015846/1$ for the *Centre for Doctoral Training in Intelligent Games and Game Intelligence* (www.iggi.org.uk) from the UK Engineering and Physical Sciences Research Council (EPSRC). For the purpose of open access, the author(s) has

Data Access Statement

All data supporting this study are provided as supplementary information accompanying [Dobre et al. 2022] at DOI:10.1007/$s10055-022-00644-4$.

REFERENCES

Jeremy Bailenson. 2018. *If a possible mass shooter wants to hone his craft, don't hand him a virtual boot camp.* https://edition.cnn.com/2018/03/05/opinions/video-games-shooting-opinion-bailenson/index.html

Dan Bohus and Eric Horvitz. 2014. Managing Human-Robot Engagement with Forecasts and... Um... Hesitations. In *Proceedings of the 16th International Conference on Multimodal Interaction* (Istanbul, Turkey) *(ICMI '14)*. Association for Computing Machinery, New York, NY, USA, 2–9. https://doi.org/10.1145/2663204.2663241

Jesper Vang Christensen, Mads Mathiesen, Joakim Have Poulsen, Ea Ehrnberg Ustrup, and Martin Kraus. 2018. Player experience in a vr and non-vr multiplayer game. In *Proceedings of the Virtual Reality International Conference-Laval Virtual*. 1–4.

Soumia Dermouche and Catherine Pelachaud. 2019. Engagement Modeling in Dyadic Interaction. In *2019 International Conference on Multimodal Interaction*. 440–445.

Svati Dhamija and Terranee E Boult. 2017. Automated mood-aware engagement prediction. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 1–8.

Georgiana Cristina Dobre, Marco Gillies, and Xueni Pan. 2022. Immersive machine learning for social attitude detection in virtual reality narrative games. *Virtual Reality* (2022), 1–20.

M. Gillies, A. Kleinsmith, and H. Brenton. 2015. Applying the CASSM framework to improving end user debugging of interactive machine learning. In *International Conference on Intelligent User Interfaces, Proceedings IUI*, Vol. 2015-Janua. https://doi.org/10.1145/2678025.2701373

Nadine Glas and Catherine Pelachaud. 2015. Definitions of engagement in human-agent interaction. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 944–949.

Goren Gordon, Samuel Spaulding, Jacqueline Kory Westlund, Jin Joo Lee, Luke Plummer, Marayna Martinez, Madhurima Das, and Cynthia Breazeal. 2016. Affective personalization of a social robot tutor for children's second language skills. In *Thirtieth AAAI Conference on Artificial Intelligence*.

Edward Twitchell Hall. 1966. *The hidden dimension*. Vol. 609. Garden City, NY: Doubleday.

Barry Ip. 2011. Narrative structures in computer and video games: Part 1: Context, definitions, and initial findings. *Games and Culture* 6, 2 (2011), 103–134.

S. Mota and R. W. Picard. 2003. Automated Posture Analysis for Detecting Learner's Interest Level. In *2003 Conference on Computer Vision and Pattern Recognition Workshop*, Vol. 5. 49–49.

Jyotirmay Sanghvi, Ginevra Castellano, Iolanda Leite, André Pereira, Peter W. McOwan, and Ana Paiva. 2011. Automatic Analysis of Affective Postures and Body Motion to Detect Engagement with a Game Companion. In *Proceedings of the 6th International Conference on Human-Robot Interaction* (Lausanne, Switzerland) *(HRI '11)*. Association for Computing Machinery, New York, NY, USA, 305–312. https://doi.org/10.1145/1957656.1957781

Albrecht Schmidt. 2000. Implicit human computer interaction through context. *Personal technologies* 4, 2 (2000), 191–199.

Mel Slater. 2009. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, 1535 (2009), 3549–3557.

Mel Slater and Anthony Steed. 2000. A virtual presence counter. *Presence: Teleoperators & Virtual Environments* 9, 5 (2000), 413–434.

Alessandro Vinciarelli, Maja Pantic, Dirk Heylen, Catherine Pelachaud, Isabella Poggi, Francesca D'Errico, and Marc Schroeder. 2011. Bridging the gap between social animal and unsocial machine: A survey of social signal processing. *IEEE Transactions on Affective Computing* 3, 1 (2011), 69–87.

Graham Wilson and Mark McGill. 2018. Violent video games in virtual reality: Re-evaluating the impact and rating of interactive experiences. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. 535–548.

Beverly Woolf, Winslow Burleson, Ivon Arroyo, Toby Dragon, David Cooper, and Rosalind Picard. 2009. Affect-aware tutors: recognising and responding to student affect. *International Journal of Learning Technology* 4, 3-4 (2009), 129–164.