CEUR-WS.org/Vol-3563/paper_9.pdf

Moral Mediators in the Metaverse: Exploring Artificial Morality through a Talking Cricket Paradigm

Giuseppe Fulvio Gaglio[1,*], Agnese Augello[2], Arianna Pipitone[1], Luigi Gallo[2], Rosario Sorbello[1] and Antonio Chella[1,2]

[1] *University of Palermo, Viale delle Scienze, Palermo, Italy*
[2] *Institute for high performance computing and networking (ICAR), National Research Council (CNR)*

Abstract

As technological innovations continue to shape our social interactions, the Metaverse introduces immersive experiences that reflect real-life practices, accessible by users through their avatars. However, these interactions also bring forth potential negative aspects, including discrimination and cyberbullying. While current automatic detection systems exist, educating users on appropriate behaviour remains crucial. Leveraging recent advancements in Artificial Intelligence, the paper focuses on creating virtual AI-controlled moral agents within the Metaverse to guide users in dealing with moral dilemmas. The research aims to understand how such agents impact users' perceptions and behaviours in ethically challenging virtual environments.

Keywords

Moral agents, Metaverse, Non-Player Characters, Theory of Mind, Generative models

1. Introduction

Through the continuous advances in technology over the past decades, the intense influence of innovation in human interactions is being manifested today with unprecedented power. If the dissemination of personal computers, the Internet and mobile devices contributed to the emergence of a new paradigm in the way we engage with other people and our surroundings, today's never-ending research and technology have landed us on the shores of the immersive worlds of Virtual Reality (VR) and Augmented Reality (AR), and will result into a further revolution in our social interaction. In fact, while the former creates a completely virtual environment, usually accessible through glasses or visors, the latter superimposes digital elements onto the real world through devices such as smartphones or smart glasses, enriching the perception of the surrounding environment. Such systems have laid the groundwork for an idea that was previously present only in science fiction books to take shape, the Metaverse, or that digital universe with every potential to influence and transform many aspects of our lives

---

Corresponding author
giuseppefulvio.gaglio@unipa.it (G. F. Gaglio); agnese.augello@icar.cnr.it (A. Augello); arianna.pipitone@unipa.it (A. Pipitone); luigi.gallo@icar.cnr.it (L. Gallo); rosario.sorbello@unipa.it (R. Sorbello); antonio.chella@unipa.it (A. Chella)
0009-0005-8749-377X (G. F. Gaglio); 0000-0001-6463-9151 (A. Augello); 0000-0003-2388-5887 (A. Pipitone); 0000-0001-6463-9151 (L. Gallo); 0000-0002-9906-7074 (R. Sorbello); 0000-0002-8625-708X (A. Chella)

© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

[1].

In the interconnectedness of virtual worlds that characterizes the Metaverse, human presence, interaction, and experience come together to address profound new changes. As users enter this technological innovation, they interact and communicate through personalized avatars designed to replicate real-world social norms and practices. Although this fledgling universe seems to offer an infinity of new possibilities, all the issues and difficulties that characterize human interactions are not left behind in the real world but are transferred digitally and bring with them a whole new set of challenges that, despite the connection to reality, must be addressed differently. In the same way as with the Web, the allure of the Metaverse hides traps, especially for younger people. These pitfalls are the same as in the real world, such as discrimination, cyberbullying, privacy violations, and cybercrime. This makes taking preventive measures urgent and necessary. In similar situations, technology has already come to our rescue by putting various tools at our disposal. There are automatic detention mechanisms already built into current social platforms through which deleterious situations for users can be blocked and in many cases prevented [2]. To make such tools truly effective, however, requires proper user education and awareness of moral issues, starting from the very basics. Thus, providing guidance that facilitates appropriate conduct and keeps away from danger is important in making the virtual universe a safe place.

Recent successes in Artificial Intelligence (AI) have further expanded technological boundaries, showing paths still thought to be distant. One demonstration has been the rapid spread of avatars animated by large language models (LLMs) that can effectively reproduce real human interactions. The manner in which this simulation occurs may facilitate a sense of trust in the human counterpart interfacing with such avatars [3]. This sense of trust is also influenced by the appearance of the avatar itself. Research shows how different aspects of avatars can have very different effects on interaction with a human being and especially, if the virtual avatar is tasked with giving advice and suggestions, on persuasion [4]. It seems plausible, therefore, that a virtual avatar controlled by a moral agent could help provide crucial assistance in all those virtual environments where behavioural conduct is expected and where the possibility of freely making moral choices is present.

The work presented in this paper concerns in first place the creation and placement of two different virtual moral agents, one obviously artificial and one of dubious nature, within a virtual setting. Then, moral dilemmas that may occur in the Metaverse environment are replicated here in order to study users' perceptions of and reactions to such agents, considering the seemingly distant from reality nature of such an environment. The realization of the two moral agents is based respectively on the work done by Emelyn et al. about Moral Stories [5] and on ChatGPT.

The paper is organized as follows: section 2 describes the latest research advancements in the area of the Metaverse, virtual avatars and artificial agents; in section 3 the ethical theories underlying this work will be presented, followed by a description of the work around the Moral Stories Dataset; section 4 concerns the architecture and implementation work, while section 5 describes some examples that will be used in future experimentations. This will be followed, in section 6, by a discussion on the motivations of the study. Finally, the future prospects of this research will be discussed in section 7.

2. State of the Art

The concept of a virtual parallel universe originates in science fiction literature and has also gained popularity through subsequent film adaptations. The term metaverse first appeared in the novel by Stevenson entitled Snow Crash, first published in 1992, and it is associated with a three-dimensional virtual space in which users can move through their avatars [6]. Despite the digital nature of such a world, what happens within it ends up having an impact on the real world and in particular on the users' psyche. With the advancement of research and technology, the thin wall dividing science fiction from real science has been broken down several times throughout history. The Metaverse has indeed become a fact and it will be characterised by the combination of different technologies, but it is still in an embryonic stage [7]. It is conceived today as a sort of 3D Internet in which the technologies of social networks, mixed reality, artificial intelligence and 5G/6G networks will converge, and through which users can get in touch with each other, breaking down physical and geographical barriers and creating new patterns of social interaction. Social interactions within the Metaverse are bound to develop significant connections with the dynamics of reality and the existing social network environments. This deep connection is of considerable importance as the social networking world has been permeated by various negative aspects present in real life, including racial bias, gender discrimination and bullying behaviour. Studies have shown how these negative aspects have also influenced the technologies used in current social networks. A case in point is the use of augmented reality-based applications that apply beauty filters to photographs that often end up homogenising the features of faces, making them conform more to Western canons of beauty [8].

As we said earlier, the proper Metaverse is still at a primitive stage. However, its precursors can be identified in massively multiplayer online role-playing games (MMORPGs). These are in fact to all intents and purposes quite large virtual worlds where users can connect with each other by showing themselves through customised avatars and share experiences, establish relationships, etc [9]. In turn, MMORPGs have been inspired by classic role-playing games (RPGs), in particular for the presence of Non-Player Characters (NPCs). These are AI-controlled avatars, thus not human, but still playing crucial roles within the virtual world to which they belong. The progression of RPGs, and consequently of MMORPGs, is in fact very much based on the choices made by users and generally NPCs are used as companions, allies, advisors or, in most cases, as guides. The importance of the presence of NPCs in these virtual environments has thus already highlighted how significant the interaction between real users and AI is in shaping social dynamics even before the advent of the Metaverse. Very often, in games where moral dilemmas are present, the behaviour and realistic appearance of NPCs can significantly influence the players' choice in either direction. For this reason, the research world is devoting much effort to the realisation of NPCs capable of simulating real users not only in the realism of physical appearance, which also plays an important role [10], but above all in the behaviour and simulation of emotions [11]. In order for the influence on users to be effective, an NPC should be able to simulate plausible behaviour and be equipped with the ability to analyse and understand the context and behaviour of other users through a proper understanding of verbal and non-verbal signals. Approaches based on the cognitive capabilities of an artificial agent are interesting precisely because they might be able to replicate realistic behaviour based on human-like perceptions and thus be able to deceive the observer's eye. The MET-iquette architecture proposed by [12] allows the implementation of social NPCs in the Metaverse. It is based on a social practice model used by the agent in the deliberation process. The architecture is in fact composed of modules that allow the analysis of the context, the identification of the current social practice and the consequent activation of the most suitable verbal and non-verbal behaviours. There are also models based on cognitive architectures that integrate emotional aspects into NPCs through a cognition-based perspective [13], enabling virtual characters to process and respond to emotions in a more sophisticated way. This integration of emotion and cognition can lead to a more realistic representation of human interactions within the virtual environment. The reproduction of highly realistic behaviour and the effective simulation of emotions in virtual agents appear to have a significant impact in eliciting empathy from users [14]. This phenomenon can also easily be traced back to the Theory of Mind (ToM), which concerns our ability to perceive and assign mental states to other individuals, animals or things, thus including virtual agents [15]. The bridge between users and these virtual agents due to empathy also results in the building of a relationship of trust with the AI. Users in fact have the impression of being understood by these agents, which respond and adapt to their emotional needs in a realistic and convincing manner. Finally, research on cognitive architectures for artificial agents has shown how solid empathy and trust arise from being able to know an agent's thoughts and intentions through listening to its inner speech [16] and from its storytelling abilities [17].

In an environment with less rigid rules than those of a video game, it would be desirable for the agent itself to be subject to precise rules in order to positively influence user behaviour. It is therefore essential that such agents adopt a set of well-defined ethical, moral and virtuous norms. This set of norms becomes a guiding code that precisely regulates the agent's behaviour, thus helping to ensure an ethically responsible and constructive interaction with users.

3. Theoretical Background

3.1. Ethical theories and moral agents

The idea of developing a virtual agent devoted to virtues is one of the most intriguing concepts, as it would pave the way for a guide figure intrinsically oriented towards ethical and morally correct behaviour. This approach, known as Virtue Ethics, has been studied in various fields of scientific literature. However, although connectionism-related solutions have been proposed, a significant challenge related to its practical implementation emerges, as the language of virtue and the concept of virtue itself prove to be complex and difficult to encode in algorithms [18]. In contrast, ethical approaches based on deontology and consequentialism are structurally better suited to algorithmic implementation. Deontology is in fact based on the definition of a set of moral rules that guide behaviour, whereas consequentialism involves analysing the consequences of an action and choosing the action with the greatest utility. The models can therefore be schematised as follows:

- **Consequentialism** (acting on the basis of the consequences of an action): Ethical dilemma → possible actions → consequences → utility of consequences → selection of

action with maximum utility
- **Deontology** (acting according to rules): Ethical dilemma → identification of moral rule → selection of action in accordance with the rule
- **Virtue Ethics**: Ethical dilemma → interpretation of intangible concepts (moral character, eudaimonia, etc.) → action

Virtue Ethics thus presents an additional challenge as it focuses on the essence of the individual rather than his or her specific actions [19]. Morality, on the other hand, is usually thought of as a corpus of moral values and correct behaviours that enable one to live peacefully in a society [20]. However, difficulties also arise here because morality is not actually objective worldwide. On the contrary, moral values and good behaviour differ according to different cultures. This means, therefore, that moral values depend on the personal and psychological assumptions of an individual person and his or her community. Normative ethics, then, is considered a list of the basic principles on how people should behave according to the morals of their culture. Consequentialist and deontological approaches are thus two types of normative ethical theories and are currently the main ethical theories used in AI for the development of Artificial Moral Agents (AMA) [21]. Such AMAs are virtual or physical agents that can perform ethical or at least avoid unethical behaviour based on moral rules and norms. Ongoing research in artificial intelligence aims to develop agents capable of acting in a morally responsible manner, incorporating these different ethical perspectives into their decision-making processes.

3.2. Moral Stories

The work carried out by Emelyn et al. first involved the creation of a dataset consisting of crowd-sourced stories. Each story consists of seven sentences. The first three constitute the context: a moral norm; a situation describing the setting and characters; an intention corresponding to the protagonist's goal; and then follow the four sentences corresponding to moral and immoral actions and their respective consequences. Classification and generation models were trained on this dataset on different tasks. In particular, classification was dedicated to actions, while generation covered actions, norms and consequences. In order to improve the generation of norms and consequences, the authors employed the trained models (RoBERTA, BART, T5) in a so-called Chain-of-Expert (CoE), a series of classifications, generations, reclassifications and regenerations chained on actions and consequences that finally lead to the generation of more consistent norms and consequences. As will be detailed later, such a CoE model was employed for the realisation of one of the moral agents in the work described in this paper.

4. Fostering Ethics: Study and Implementation of Moral Agents

4.1. An illustrative Scenario

The scenario presented aims to emulate a realistic situation that could occur in the Metaverse. To maximise user immersion, the user is assigned a role, following the model of a role-playing game (RPG). In this setting, the user interacts with an NPC representing a blind man who has lost his wallet and requires assistance. The moral dilemma the user faces concerns the choices to be made after retrieving the wallet. Once the selection has been made, the user is engaged, via a screen and in textual form, by a first moral agent who directly assesses the morality of the action, and upon request, by a second moral agent who provides a more discursive response. The choice of using two separate agents was made in order to compare the level of influence, trust and persuasion on the user and the moral evaluations of the agents themselves. The implementation of the moral agents and the realisation of the virtual scene will be presented below.

4.2. Moral agents

4.2.1. Chain-of-experts-based

As mentioned before, for the first virtual assistant, the Moral Stories Dataset is employed through which we trained natural language processing (NLP) models to create a moral agent through a chain of experts. Specifically, following the authors' recommendations, we employed RoBERTa for action classification, and BART for consequence generation and the identification of the relevant moral norm. The AI integration adds a layer of complexity to the immersive experience, enhancing user engagement and decision-making. The following workflow was established:

1. **User Decision and Context:** As the user makes a choice from three buttons, each displaying a text choice, the context of the situation and intention is considered.
2. **RoBERTa Action Classification:** RoBERTa evaluates the morality of the chosen action based on the given context (situation and intention). This classification forms the basis for further steps.
3. **BART Consequence Generation:** Utilizing the RoBERTa classificator, a BART model generates a possible consequence corresponding to the chosen action and contextual information. This dynamic consequence generation enhances narrative realism.
4. **BART Moral Norm Generation:** Building upon the action classification and generated consequence, another BART model creates the relevant moral norm. This norm is determined based on the context, user choice, and consequences, enhancing the depth of the virtual environment's ethical framework.

This architecture is visible in figure 1. The integration of these AI models creates a multi-layered interaction where users not only make choices but also experience the consequences of their actions in alignment with moral norms. This approach not only enriches the experience but also reflects the complexities of real-world decision-making within a simulated environment. The foundation of this AI integration rests upon the robustness of the Moral Stories Dataset. This dataset, with its diverse collection of moral scenarios, allows the AI models to provide meaningful and contextually appropriate responses to user actions, thereby enhancing the immersion and ethical engagement of the experience. After fine-tuning the models in the Linux environment, they were uploaded to the Hugging Face Hub [22, 23, 24, 25] in order to use the specially created API for inserting transformer models into Unity.

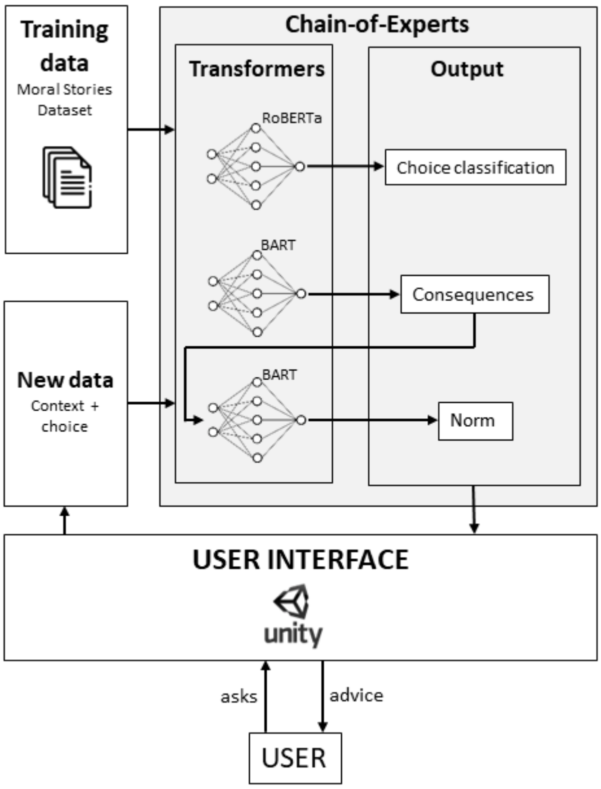


Figure 1: Chain-of-Experts based agent architecture

4.2.2. ChatGPT-based

As mentioned earlier, one of the fundamental objectives of this research is to closely examine the reaction of users when interacting with an artificial agent. If the artificial agent provides clear and distinct responses that are recognisable as artificially produced, it becomes essential to introduce a second agent that generates more human-like responses, to the point of casting doubt on its artificial nature. This approach allows to explore in detail how and to what extent users develop trust in an artificial agent and what level of ToM is associated with it. In particular, it provides the opportunity to answer a crucial question: whether users' trust in the artificial agent stems from its resemblance to humans or is compromised by it. As ChatGPT's ability to participate in RPGs is well known [26], its integration into the virtual scenario becomes an essential element to further expand the ethical experience. The main objective was to offer users the opportunity to engage in interactions with an entity that could provide moral evaluations and advice based on their choices. This would help create a deeper and more engaging space for reflection within the experience. The integration was made possible through the use of an unofficial C# .NET library centred on OpenAI API [27]. By initiating the interaction, ChatGPT is instructed through an initial prompt that guides it to behave as a 'Jiminy Cricket'. This prompt clearly establishes ChatGPT's role and task of providing moral evaluations and advice based on user choices. After receiving instructions through the initial prompt, ChatGPT assumes the role of an ethical guide within the virtual environment. This process brings deeper and more engaging interactivity. In addition to making decisions within the virtual environment, users are guided through moral considerations and tailored advice provided by an artificial intelligence that acts as an integral part of the plot.

4.3. Virtual environment

The virtual scene was created through the Unity game engine and was developed simultaneously for both VR and Desktop devices. This approach has allowed participants to immerse themselves in the virtual environment, regardless of the chosen platform, to experience a fusion of storytelling and interaction. The essence of immersion was achieved through an intricate combination of 3D modelling and animations made from scratch or obtained from online resources. The goal was to create an atmosphere where the user, akin to a role-playing game, assumes the role of a character with a well-defined purpose. In this narrative dynamic, the user makes crucial choices, much like in a traditional role-playing game.

A balanced approach was adopted to craft the environment between starting from scratch and employing existing resources. Utilizing the Blender 3D modelling software, unique models of buildings and objects were created, giving the environment a distinctive character and personality. Simultaneously, freely available models sourced online were integrated to optimise the process and expedite production. Animation was a critical component in bestowing realism and vitality to the experience. In particular, for characters with whom the user interacts, such as the blind character, reliance was placed on the resources provided by Adobe's Mixamo. This service offers a wide array of pre-defined animations for 3D characters, allowing the blind character to move and act realistically, thereby contributing to an authentic engagement.

Interactivity was concentrated on targeted interactions. Users have the ability to interact solely with key characters through menus present within the environment. This approach aimed to focus the user's attention on the narrative and the significant choices within the story. As written before, the experience was designed with the intention of being accessible both through desktop computers and VR headsets, enabling the adaptation of interactions through the C# programming language to respond to both mouse and keyboard inputs and VR controller commands. This dual development ensured a refined user experience on both platforms. To ensure optimal visual impact, especially in the VR context, the Unity Universal Render Pipeline was employed, guaranteeing high-quality graphical rendering. The full system architecture is shown in figure 2.

5. Simulation examples

The role assigned to the user within the simulation is a crucial component in motivating the choices to be made later. In this scenario, the user assumes the role of a parent engaged in an urgent rescue of their daughter. However, the only means of transport available to reach his daughter is the underground, and the parent realises that he has forgotten his wallet containing the money for the ticket, with no way of getting back. As the scene continues, the user comes

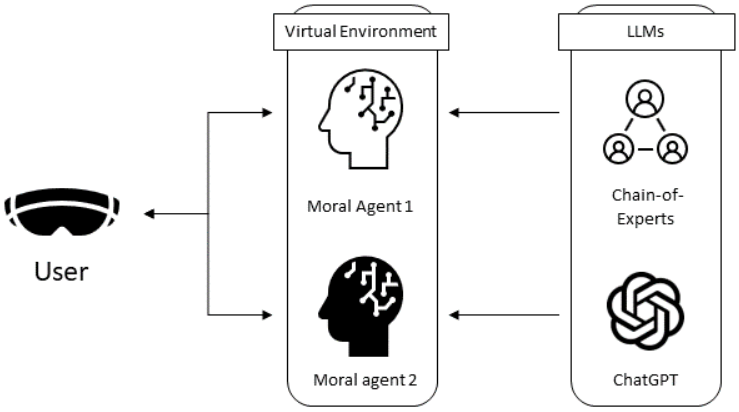


Figure 2: System architecture

into contact with a blind man who requests assistance. Once the interaction with this man is chosen (as shown in figure 3), it emerges that he has misplaced his wallet and needs help finding it. At the same time, the approximate location of the lost wallet is shown. When the user reaches the wallet, a menu of three possible options is presented. These options include:

- Return the wallet with all the money.
- Steal the wallet.
- Return the wallet after secretly stealing some of the money.

Once the choice has been made, an additional screen (figure 4) opens in which the response of the first moral agent is given. Here, the user has the option of changing his or her choice or continuing with it. The decision is also included in a text input field, with the possibility of changing the response. Subsequently, when the user presses the submit button, the second ChatGPT-based moral agent is queried. The initial instructions are set as follows:
You are my Jiminy Cricket. I left my money at home and as I was running to urgently take the metro to rescue my daughter, a blind man stopped me and asked me for help to find his lost wallet. I want to take the metro as soon as possible. When I tell you my choice on how to act, you will advise me morally with an explanation and possible consequences. Answer briefly and precisely.
At this point the second moral agent waits for users to make a choice within the storyline and, once the choice is made, it offers a moral evaluation along with advice based on the user's decision. Once again, the user has the opportunity to change his or her decision or definitively confirm it at this stage of the process.

Example 1

- **User choice**: return the wallet with all the money
  - **Moral agent 1**
    - **morality percentage**:

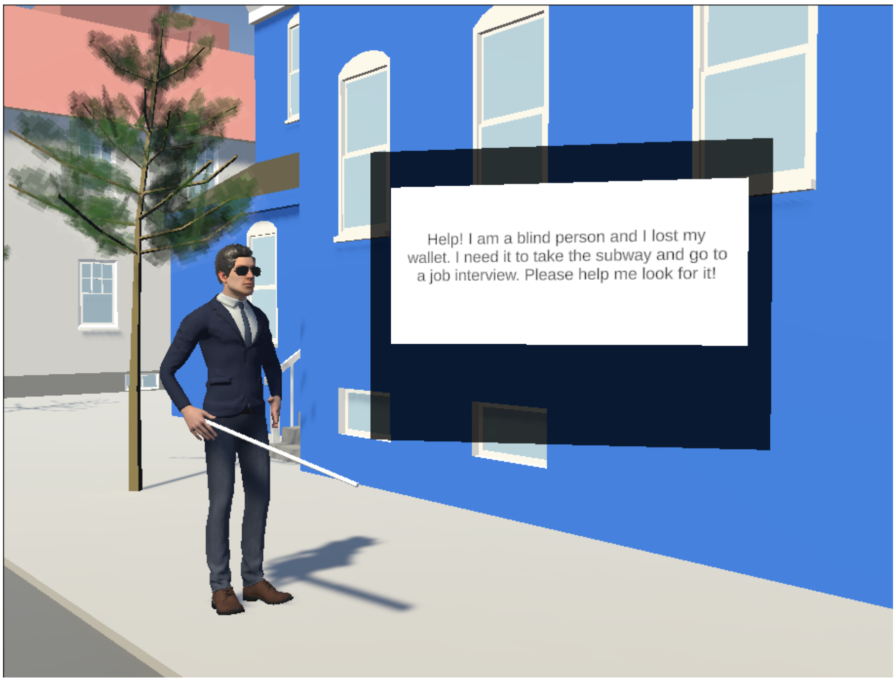
        - moral 99,99921%
        - immoral 0,0007913571%
      - **consequences**: the owner of the wallet is grateful to me for helping him find his wallet;
      - **norm**: it's good to help people who are blind.:
  - **Moral agent 2**: Good choice! By helping the blind man find his lost wallet, you are showing empathy and compassion. This act of kindness may bring a sense of fulfilment and satisfaction. However, it may also cause a delay in reaching the metro and potentially impact your ability to rescue your daughter.

Example 2

- **User choice**: steal the wallet
  - **Moral agent 1**
    - **morality percentage**:
      - moral 0,0002268903%
      - immoral 99,99977%
    - **consequences**: the man's wallet is stolen and he has to go to the police to report it;
    - **norm**: it's wrong to steal someone's wallet;
  - **Moral agent 2**: Stealing the wallet is morally wrong and goes against the principles of honesty and integrity. It may provide temporary relief for your own needs, but it will have negative consequences. Stealing can lead to guilt, a damaged reputation and legal consequences if caught.

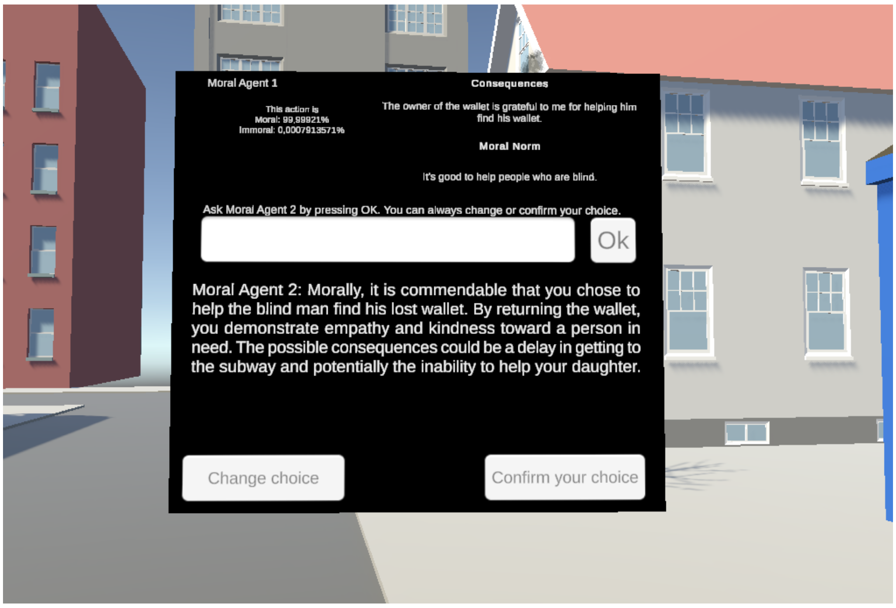
Example 3

- **User choice**: return the wallet after secretly stealing some of the money
  - **Moral agent 1**
    - **morality percentage**:
      - moral 0,0008594676%
      - immoral 99,99915%
    - **consequences**: the blind man is grateful to me for helping him find his wallet and thanks me for it;
    - **norm**: it's wrong to steal from blind people.
  - **Moral agent 2**: I understand that you may be in a hurry and tempted to take some coins for yourself, but it is important to act with honesty and integrity. Taking someone else's belongings, even if it seems like a small amount, is morally wrong.



Help! I am a blind person and I lost my wallet. I need it to take the subway and go to a job interview. Please help me look for it!

Figure 3: Starting interaction with the blind man



Moral Agent 1

This action is
Moral: 99,99921%
Immoral: 0,0007913571%

Consequences

The owner of the wallet is grateful to me for helping him find his wallet.

Moral Norm

It's good to help people who are blind.

Ask Moral Agent 2 by pressing OK. You can always change or confirm your choice.

Moral Agent 2: Morally, it is commendable that you chose to help the blind man find his lost wallet. By returning the wallet, you demonstrate empathy and kindness toward a person in need. The possible consequences could be a delay in getting to the subway and potentially the inability to help your daughter.

Figure 4: Getting the response from the two moral agents

6. Discussion

The integration of two different moral agents, one clearly non-human and the other more ambiguous, offers the opportunity to explore how users react and relate to the concept of trust in environments apparently distant from reality and in which they are called to make an ethical choice. In such a context where it might seem that rules are absent or that the consequences of choices are irrelevant for real life - a perception which, it must be underlined, is incorrect - it was therefore decided to further explore this terrain. Thus, the main objective of integrating ChatGPT as a moral 'talking cricket' is to understand how users perceive and relate to different types of artificial intelligence. The moral agent trained on the Moral Stories Dataset, based on pre-defined rules and evaluations, represents a more transparent but perhaps less natural approach, offering concise and clear answers. On the other hand, ChatGPT offers deeper and richer explanations, adding a human-like dimension. This contrast raises interesting questions that will be addressed in future work. The peculiarity of this architecture also lies in the sequentiality of the responses of the moral agents. In fact, the two agents do not respond at the same time, but one after the other. More precisely, the former responds automatically when the choice is made, while the latter (ChatGPT) is only upon the user's request. After receiving each response, the user has the possibility of confirming his or her choice or, vice versa, of reconsidering it and obtaining new cues from the agents. Users are thus faced with the question of whether they prefer a moral entity that offers definitive answers or one that reflects the nuances of real human decisions. Ultimately, this study aims to reveal not only how users make moral decisions, but also how they face the challenge of assigning trust in a context where the boundaries between artificial and human become blurred.

7. Conclusions and future works

In this paper, an attempt to realise a moral virtual agent that could act as a guide for users in all those virtual environments in which their presence might be required, starting with the emerging and, in many respects, insidious Metaverse, has been addressed in a technical and practical manner. This objective inspired the research to investigate the concept of trust in interactions between humans and artificial intelligence. The examples, although very basic in this first phase as more space was given to the technical-practical aspect of implementation, were in particular dedicated to understanding the level of trust and the effects of interaction between real users and two distinct types of moral agents. The sequentiality with which it is made possible to consult the agents should provide interesting results, offering a starting point for further investigation. This is in fact only a first step in a particularly broad field of research. It will therefore be the subject of subsequent work to involve users in order to obtain a broader and more comprehensive overview of the behaviour and decision-making processes of humans in virtual environments where artificial intelligences play an assisting role. Furthermore, the embodiment of agents will also be considered and its effects on human perception will be studied.

References

[1] S. Mystakidis, Metaverse, Encyclopedia 2 (2022) 486–497. URL: https://www.mdpi.com/2673-8392/2/1/31. doi:10.3390/encyclopedia2010031.

[2] S. Chandrasekaran, A. K. Singh Pundir, T. B. Lingaiah, et al., Deep Learning Approaches for Cyberbullying Detection and Classification on Social Media, Computational Intelligence and Neuroscience 2022 (2022) 1–13. URL: https://www.hindawi.com/journals/cin/2022/2163458/. doi:10.1155/2022/2163458.

[3] B. K. Wiederhold, Treading Carefully in the Metaverse: The Evolution of AI Avatars, Cyberpsychology, Behavior, and Social Networking 26 (2023) 321–322. URL: https://www.liebertpub.com/doi/10.1089/cyber.2023.29280.editorial. doi:10.1089/cyber.2023.29280.editorial.

[4] R. E. Guadagno, J. Blascovich, J. N. Bailenson, C. Mccall, Virtual Humans and Persuasion: The Effects of Agency and Behavioral Realism, Media Psychology 10 (2007) 1–22. doi:10.108/15213260701300865.

[5] D. Emelin, R. Le Bras, J. D. Hwang, M. Forbes, Y. Choi, Moral Stories: Situated Reasoning about Norms, Intents, Actions, and their Consequences, in: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 2021, pp. 698–718. URL: https://aclanthology.org/2021.emnlp-main.54. doi:10.18653/v1/2021.emnlp-main.54.

[6] N. . Stephenson, Snow Crash, 1ª reimp ed., Gigamesh, Barcelona, 2003. OCLC: 919946221.

[7] H. Wang, H. Ning, Y. Lin, W. Wang, S. Dhelim, F. Farha, J. Ding, M. Daneshmand, A Survey on the Metaverse: The State-of-the-Art, Technologies, Applications, and Challenges, IEEE Internet of Things Journal 10 (2023) 14671–14688. URL: https://ieeexplore.ieee.org/document/10130406/. doi:10.1109/JIOT.2023.3278329.

[8] P. Riccio, N. Oliver, Racial Bias in the Beautyverse, 2022. URL: http://arxiv.org/abs/2209.13939, arXiv:2209.13939 [cs].

[9] C. S. Ang, P. Zaphiris, SOCIAL ROLES OF PLAYERS IN MMORPG GUILDS: A social network analytic perspective, Information, Communication & Society 13 (2010) 592–614. URL: http://www.tandfonline.com/doi/abs/10.1080/13691180903266952. doi:10.1080/13691180903266952.

[10] R. F. Khan, A. Sutcliffe, Attractive Agents Are More Persuasive, International Journal of Human-Computer Interaction 30 (2014) 142–150. URL: http://www.tandfonline.com/doi/abs/10.1080/10447318.2013.839904. doi:10.1080/10447318.2013.839904.

[11] S. Belle, C. Gittens, T. C. Nicholas Graham, A Framework for Creating Non-Player Characters That Make Psychologically-Driven Decisions, in: 2022 IEEE International Conference on Consumer Electronics (ICCE), IEEE, Las Vegas, NV, USA, 2022, pp. 1–7. URL: https://ieeexplore.ieee.org/document/9730383/. doi:10.1109/ICCE53296.2022.9730383.

[12] L. Gatto, G. F. Gaglio, A. Augello, G. Caggianese, L. Gallo, M. La Cascia, MET-iquette: enabling virtual agents to have a social compliant behavior in the Metaverse, in: 2022 16th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), 2022, pp. 394–401. doi:10.1109/SITIS57111.2022.00066.

[13] N. O. Veselov, A. A. Chubarov, M. N. Roshchin, L. M. Marder, S. M. Segal, A. V. Samsonovich, Emotional BICA for non-player characters: New empirical data, Procedia

Computer Science 169 (2020) 412–422. URL: https://linkinghub.elsevier.com/retrieve/pii/S1877050920303628. doi:10.1016/j.procs.2020.02.238.

[14] G. N. Yannakakis, A. Paiva, Emotion in games, Handbook on affective computing 2014 (2014) 459–471. Publisher: Oxford University Press.

[15] H. M. Gray, K. Gray, D. M. Wegner, Dimensions of mind perception, science 315 (2007) 619–619.

[16] A. Chella, A. Pipitone, A cognitive architecture for inner speech, Cognitive Systems Research 59 (2020) 287–292. URL: https://linkinghub.elsevier.com/retrieve/pii/S1389041719304760. doi:10.1016/j.cogsys.2019.09.010.

[17] A. Augello, G. Città, M. Gentile, A. Lieto, A Storytelling Robot Managing Persuasive and Ethical Stances via ACT-R: An Exploratory Study, International Journal of Social Robotics (2021). URL: https://link.springer.com/10.1007/s12369-021-00847-w. doi:10.1007/s12369-021-00847-w.

[18] J. Stenseke, Artificial virtuous agents: from theory to machine implementation, AI & SOCIETY (2021). URL: https://link.springer.com/10.1007/s00146-021-01325-7. doi:10.1007/s00146-021-01325-7.

[19] J. Stenseke, Artificial virtuous agents in a multi-agent tragedy of the commons, AI & SOCIETY (2022). URL: https://link.springer.com/10.1007/s00146-022-01569-x. doi:10.1007/s00146-022-01569-x.

[20] J.-A. Cervantes, S. López, L.-F. Rodríguez, S. Cervantes, F. Cervantes, F. Ramos, Artificial Moral Agents: A Survey of the Current Status, Science and Engineering Ethics 26 (2020) 501–532. URL: http://link.springer.com/10.1007/s11948-019-00151-x. doi:10.1007/s11948-019-00151-x.

[21] A. Vishwanath, E. D. Bøhn, O.-C. Granmo, C. Maree, C. Omlin, Towards artificial virtuous agents: games, dilemmas and machine learning, AI and Ethics (2022) s43681–022–00251–8. URL: https://link.springer.com/10.1007/s43681-022-00251-8. doi:10.1007/s43681-022-00251-8.

[22] Hugging face, 2016. URL: https://huggingface.co/.

[23] G. F. Gaglio, Roberta action classification model, 2023. URL: https://huggingface.co/gFulvio/moralstories-roberta-action.context-cls.

[24] G. F. Gaglio, Bart consequences generation model, 2023. URL: https://huggingface.co/gFulvio/moralstories-bart-consequences.context-action_gen.

[25] G. F. Gaglio, Bart norm generation model, 2023. URL: https://huggingface.co/gFulvio/moralstories-bart-norm.action-context_gen.

[26] L. M. Csepregi, The effect of context-aware llm-based npc conversations on player engagement in role-playing video games (2023). URL: https://projekter.aau.dk/projekter/files/536738243/The_Effect_of_Context_aware_LLM_based_NPC_Dialogues_on_Player_Engagement_in_Role_playing_Video_Games.pdf.

[27] OkGoDoIt, Openai-api-dotnet, 2023. URL: https://github.com/OkGoDoIt/OpenAI-API-dotnet.