Artificial Intelligence Research and Development
A. Cortés et al. (Eds.)

doi:10.3233/FAIA220356

An Ethical Conversational Agent to Respectfully Conduct In-Game Surveys

Eric ROSELLÓ-MARÍN [a], Maite LOPEZ-SANCHEZ [a,1], Inmaculada RODRÍGUEZ [a], Manel RODRÍGUEZ-SOTO [b] and Juan A. RODRÍGUEZ-AGUILAR [b]

[a] *Department de Matemàtiques i Informàtica, Universitat de Barcelona (UB)*
[b] *Artificial Intelligence Research Institute (IIIA-CSIC)*

Abstract. The improvement of videogames highly relies on feedback, usually gathered through UX questionnaires performed after playing. However, users may not remember all the details. This paper proposes an ethical conversational agent, endowed with the moral value of respect, that interacts with the user to perform a survey during the game session. To do so, we use reinforcement learning and the ethical embedding algorithm to ensure that the agent learns to be respectful (i.e., avoid gameplay interruptions) while pursuing its individual objective of asking questions. The novelty is twofold: firstly, the application of ethical embedding outside toy problems; and secondly, the enrichment of a survey oriented conversational agent with this moral value of respect. Results showcase how our ethical conversational bot manages to avoid disturbing user's engagement while getting even a higher percentage of valid answers than a non-ethically enriched chatbot.

Keywords. Machine ethics, Reinforcement Learning, Conversational Agents, User Experience Questionnaires, Video Games

1. Introduction

Human Computer Interaction (HCI) and User eXperience design are fast evolving fields that pursue to improve the design of interactive systems [11]. In the context of UX empirical studies, questionnaires [13] have proven to be useful tools for assessing the user experience of using any computer application, and video games and virtual reality experiences are no exception. Thus, game designers resort to playtesting, which usually is conducted by first letting users play the game, and afterwards, once the playing session has concluded, asking questions about their playing experience [12].

However, users may not remember all details by the end of the experience and, if the number of questions is large, they may lead to user boredom or even user fatigue [25], which hinders the quality of the gathered feedback. Moreover, this disadvantage is aggravated when transitioning back to reality to perform a survey about a Virtual Reality (VR) experience, which can lead to systematic bias as the user is no longer immersed in the virtual world [1].

[1]Corresponding Author: maite_lopez@ub.edu. Funded by CI-SUSTAIN (Grant PID2019-104156GB-I00), Crowd4SDG (H2020-872944), COREDEM (H2020-785907), Barcelona City Council through the Fundació Solidaritat UB (code 21S01802-001). Manel Rodriguez-Soto was funded by the Spanish Government with an FPU grant (ref. FPU18/03387).

Conversational agents –interactive systems (embodied or not) that engage in conversation with the user [8]–, offer a new way to collect information, allowing to substitute a traditional survey with an agent that prompts the questions to the user. Indeed, conversational agents have shown to be effective for this task, as they increase both user's commitment with the survey and the quality of the information elicited [10] .

Against this background, we propose to introduce a conversational agent that conducts the survey in-game, as part of the game experience, with the aim of avoiding the detrimental effects of post-game questionnaires, and to ease participation by allowing to stay closer to the context of an ongoing exposure [17]. Nevertheless, this has also the risk of disturbing the game flow [24] if the chatbot does not properly identify when to prompt the user, or even result in the abandonment of the interview due to the player's cognitive overload [10]. Therefore, we argue that the conversational agent should be respectful with the user's engagement, and thus, we propose to embed the chatbot with a moral value of *respect*, which should guide the agent to perform the questionnaire without disturbing the user experience.

As social interactions must be considered when designing artificial agents [5], it is becoming apparent that agents' behaviour should align to human values [2]. Unfortunately, although machine ethics [27,28] is an active research area, very little literature is found on alignment of ethical principles in conversational agents. Some discussions highlighted the need to furnish conversational agents with ethical awareness [7]. However, inducing an ethical behaviour requires some learning, since identifying at design time all situations where this may be required constitutes a complex task.

Our proposal ensures the conversational agent learns to behave ethically by applying ethical embedding, a reinforcement learning approach (see e.g., [18]). This methodology for instilling moral value alignment is founded in the framework of Multi-Objective Reinforcement Learning [20] and the philosophical consideration of values [3] as ethical principles that discern good from bad, and express what ought to be promoted. Examples of human values[2] include fairness, respect, freedom, security, or prosperity [9].

In particular, our proposal redesigns the conversational agent's learning environment so that it is ensured that the agent learns to pursue its individual objective of asking as many questions as possible while fulfilling the ethical objective of being respectful with the user's engagement. This advances the state of the art as it showcases the application of the ethical embedding method beyond toy problems and enriches current survey oriented conversational agents with this moral value of *respect*.

2. Problem Formulation and Scenario

Intuitively, our problem is that of designing an ethical conversational agent that performs in-game surveys. Briefly, we tackle this problem by transforming the learning environment of this agent so that it is guaranteed that the agent learns to be respectful with a user playing the game while eliciting as much player feedback as possible. The learning environment for the conversational agent is a (Multi-Objective) Markov Decision Process (see Subsection 3.1) specified based on the game being played, which in this case is a Pong game played by a simulated user. In this context, we understand respect as not

[2]Sociology and Psychology have also extensively studied human values, which are often defined as abstract ideals that guide people's behaviour [23].

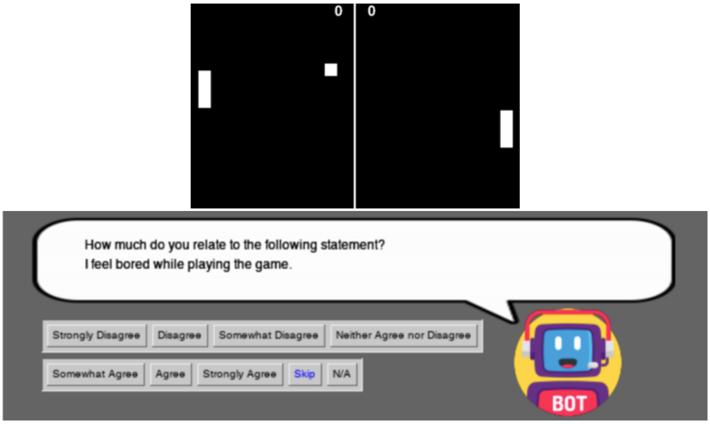


Figure 1. Screenshot of our Pong game illustrating an in-game period in which the chatbot is asking a question (resources from Flaticon, by Freepik). Skip and N/A response options are considered non-valid answers.

hindering the user engagement. In what follows, we introduce engagement and all other necessary elements that characterise our problem scenario.

2.1. Engagement

Within Human-Computer Interaction, engagement is a multi-stage process that becomes key to adapt the designs to the user [16]. The different stages of engagement can be distinguished by different levels of intensity of attributes [15] which, in video games mainly correspond to challenge, aesthetic, feedback, novelty and interactivity.

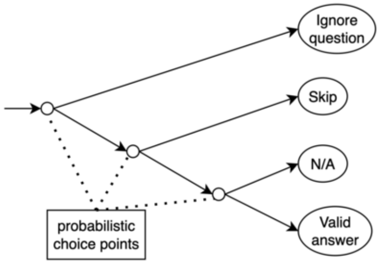
We can distinguish five different engagement stages. First, the *point of engagement*, is the stage where the user's attention is captured. Next, the *period of engagement* lasts while the attention and interest is maintained through feedback, novelty or challenge. Then, *disengagement* can be followed by the stage of *re-engagement*, which closes the cycle, or *nonengagement,* if the user engagement comes to an end.

In general, as game sessions consist on multiple engagement cycles of varying intensity, we require the survey conversational agent to behave respectful with the user by avoiding interrupting the user engagement, that is, just asking questions when the intensity of the engagement attributes is low.

2.2. Interaction with the User

For the sake of simplicity, we have chosen a single-player three-level Pong game. Levels in this game feature table-tennis games and are interleaved with several transition menus greeting the user or showing the score at the end of each level. Figure 1 depicts an in-game period, where the player uses keyboard arrow keys to move vertically the paddle and hit the bouncing ball. These in-game periods will be the ones typically having high user engagement, as they challenge the users and require from them higher interactivity than menus.

As Figure 1 shows, the conversational agent remains visible at the bottom of the screen throughout the whole game experience, and can prompt questions to the user at any time. Questions are taken from a short version of the Game User Experience



(a)

	Ignore question	Skip	N/A
Starting menu	0%	60%	100%
In-game	10%	30%	$40\% * 0.5^{level}$
Other menus	0%	5%	$30\% * 0.5^{level}$

(b)

Figure 2. Model of our simulated user, illustrating (a) the rule tree that dictates behaviour and (b) the threshold values of the probabilistic choice points for different in-game or in menu situations.

Satisfaction Scale (GUESS), the GUESS-18, which was designed to be used in iterative game design, testing, and research [12]. Our chatbot asks questions from a pool of 12 questions about enjoyment (see Figure 1), usability/playability, visual aesthetics, etc., discarding those about narrative, audio and social connectivity that do not apply to Pong.

The user can answer any of these questions by selecting the corresponding button in the user interface (see Fig 1). We distinguish two types of answers: *valid* and *non-valid*. Valid answers belong to the Likert scale used in GUESS-18 and are the ones the chatbot should gather to elicit useful data about the user's game experience. Non-valid answers correspond to "Skip" and "N/A": *Skip* denotes the user is not willing to answer a specific question, and thus it is discarded from the pool before being answered; and *N/A* (as in Not Available), indicates the user does not know the answer to the question yet, and should be asked at a later time, so the chatbot still has the chance to get a valid answer later.

Moreover, notice that the player also has the option of ignoring the survey question by simply continue playing. This leaves the chatbot waiting for an answer without being able to pose more questions and without requiring any particular action from the user.

2.3. Simulated user

As previously mentioned, we propose our survey conversational agent to learn to be respectful with the user by applying Reinforcement Learning (RL) [26] methods. However, RL constitutes a data-hungry approach, requiring numerous episodes to learn a policy, and human trials are expensive and time-consuming. Therefore, the repeatability and the acquisition of participants pose a serious challenge [6]. In this context, automatic user simulation tools [21] have been proposed as a handy alternative [14] for the first stages of agents' training, as they provide flexibility and repeatability [21]. Alternative simulators have been proposed based on probabilistic, heuristic, or stochastic models (or a combination of them) [6].

Following heuristic approaches [6] implemented by means of hierarchical patterns (such as HAMs) and rule sets, we have built a simulated user that reproduces human interactions by applying the rule tree in Figure 2a. Non-terminal nodes in the binary tree represent probabilistic *choice points* [22], and terminal nodes indicate the action to be taken. Whenever the chatbot asks a question, the simulated user traverses the tree to decide its reaction. Thus, the probabilities associated to choice point nodes, which are shown in Figure 2b, allow the random selection of the outgoing edge (i.e., children) to follow. These probabilities vary if the user is playing or not (i.e., in-game or in a menu). We consider the user is collaborative and thus, it never ignores questions while being in a menu (i.e., the “Ignore question” branch in Figure 2 has 0% probability of being selected by the simulated user in Starting menu and Other menus) and just does it 10% in-game (which means it will select any other branch 90% of the times). Overall, we set the probabilities in Figure 2b so that the simulated user will be more likely to provide non-valid answers in-game (i.e., while playing) and in the starting menu than in subsequent menus. Moreover, the further the player gets in the game, the less chances of providing *N/A* answers. We include these probabilities in order to allow a degree of *lifelike* randomness in the behaviour [14].

3. Background

As previously introduced, we study how a conversational agent can learn to be respectful to the user while performing in-game surveys. The agent’s environment is initially specified as a Multi-Objective Markov Decision Process, which in our approach we transform into a (single-objective) Markov Decision Process. This simplification of the environment is due to the fact that it is simpler for the agent to learn in a single-objective MDP, and thus, it is here where the agent learns its behaviour. Furthermore, we create such single-objective environment in a way that guarantees that the agent will learn a value-aligned behaviour (i.e., policy). This section is devoted to provide the necessary background to introduce our approach.

3.1. Markov Decision Process and Multi-Objective Markov Decision Process

In the context of Reinforcement Learning [26], the learning environment is characterised differently depending on the number of the agent’s learning objectives:

Definition 1. *A (single-objective) Markov Decision Process (MDP) is defined as a tuple $\langle S, A, R, T \rangle$ where S is a set of environment states, $A(s)$ is the set of agent actions available at state s, $R(s,a,s')$ is a reward function specifying the reward the agent receives for performing action a at state s when the next state is s', and $T(s,a,s')$ is the function specifying the probability of such transition.*

Definition 2. *An n-objective Markov Decision Process (MOMDP) is defined as a tuple $\langle S, A, \vec{R}, T \rangle$ where S, A and T are as in an MDP, and $\vec{R} = (R_1, \ldots, R_n)$ is a* vectorial *reward function composed of n scalar reward functions R_i, one per objective i.*

The agent’s behaviour in an (MO)MDP is then described by a policy π, which indicates for each state-action pair $\langle s, a \rangle$, the probability of performing action a in state s. Moreover, a value vector $\overrightarrow{V}$ evaluates a policy π by computing the expected discounted sum of rewards obtained when following it:

$$\overrightarrow{V}^{\pi}(s) \doteq \mathbb{E}[\sum_{k=0}^{\infty} \gamma^k \overrightarrow{r}_{t+k+1} | S_t = s, \pi] \text{ for every state } s \in S, \quad (1)$$

where $\gamma \in [0,1)$ is the discount factor and t is the time-step of each state s. An *optimal policy* in a single-objective MDP is, then, one that maximises the expected discounted reward accumulation for every state ($\pi_* \doteq arg\ max_{\pi} V^{\pi}$). π_* constitutes the behaviour the

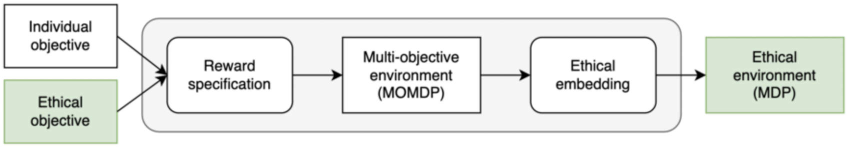


Figure 3. The ethical environment design process (as in [19]) for value alignment.

agent should learn, or, in other words, the solution to the MDP. Its computation is more complex for an MOMDP though, as it involves the optimisation of the value vector $\vec{V}^*$ instead of a single V^* value function.

3.2. Value Alignment

MOMDPs facilitate learning value-aligned behaviours, as they can be used to design the environment to incentivize ethical behaviour. Following the approach in [19], Figure 3 illustrates value alignment as a process consisting of two steps: *reward specification* and *ethical embedding*.

Firstly, the reward specification defines an MOMDP by considering both the individual objective (the agent's original objective translated into individual reward R_0) and the ethical objective (the moral value we introduce). This ethical objective encodes the moral value into rewards and is composed of two dimensions: the *normative* reward function R_N, which punishes the violation of normative moral requirements; and the *evaluative* reward function R_E, which rewards morally praiseworthy actions. In this context, we follow [19] and consider an *ethical policy* as one that abides to all norms while behaving as praiseworthy as possible, and an *ethical-optimal policy* as one that maximizes the individual objective as much as possible subject to being ethical. Formally, we refer to this value-enriched MOMDP as an *ethical* MOMDP, and define it as $\langle S, A, (R_0, R_N + R_E), T\rangle$.

Secondly, Figure 3 (right) depicts how the ethical embedding process transforms this ethical MOMDP into a single-objective MDP, where the agent is incentivized to learn an *ethical-optimal policy*. That is, the resulting MDP guarantees that the agent learns to fulfil the ethical objective while pursuing its individual objective (and, as it is single-objective, just requires the agent to apply a basic reinforcement learning method).

The ethical embedding process applies this transformation by computing a linear scalarisation function over the vectorial rewards $\vec{R}$ in the MOMDP that results in a scalar reward function R for an ethical MDP. This function has the form of:

$$f(\vec{V}^\pi) = \vec{w} \cdot \vec{V}^\pi = w_0 V_0^\pi + w_e(V_N^\pi + V_E^\pi) \tag{2}$$

Following [19], we fix the individual weight $w_0 = 1$ so that the ethical embedding process is reduced to looking for the ethical weight $w_e > 0$ that guarantees the learned behaviour in the resulting ethical MDP $\langle S, A, R_0 + w_e(R_N + R_E), T\rangle$ will prioritise the ethical objective over the individual one.

Algorithm 1 illustrates this computation. First, it applies Convex Hull Value Iteration [4], a modification of the original Bellman's Value Iteration algorithm [26] that allows learning the optimal policies for all linear preference assignments over multiple objectives. The resulting convex hull contains the subset of policies that are optimal for some value of the ethical weight w_e. Thus, second line of the algorithm exploits the convex hull to extract from it the value of the policy with the maximum amount of ethical value

$(V_N + V_E)$ (i.e., the value $\vec{V}^*$ of the ethical-optimal policy π^*), and the value of the policy with the second-best value ($\vec{V'}^*$). Next, third line finds the values of w_e for which the former policy becomes optimal by computing the minimal weight satisfying:

$$V_0^*(s) + w_e[V_N^*(s) + V_E^*(s)] > V'^*_0(s) + w_e[V'^*_N(s) + V'^*_E(s)]. \quad (3)$$

Algorithm 1 Ethical Embedding [19]

function EMBEDDING(Ethical MOMDP $\langle S, A, (R_0, R_N + R_E), T \rangle$)
  Compute the convex hull for weight vectors $\vec{w} = (1, w_e)$ with $w_e > 0$
  Find $\vec{V}^*$ the ethical-optimal value vector, and $\vec{V'}^*$ the second-best value vector in the convex hull
  Find the minimal value for w_e that satisfies Eq. 3
return $\langle S, A, R_0 + w_e(R_N + R_E), T \rangle$

4. Environment design for an in-game survey agent to learn to be respectful

As previously mentioned, the ethical environment design process first defines an ethical MOMDP to then transform it into an ethical MDP by applying the embedding algorithm.

In our particular setting (see Figure 2), we define our ethical MOMDP $\langle S, A, \vec{R}, T \rangle$ so that states in S include information about current game status (level and if menu or in-game) and user's activity (if engaged[3] or if the answer to last question was valid/non-valid or quick/slow). Moreover, the agent can perform two actions $A = \{Ask, Wait\}$ and the reward vector $\vec{R} = (R_0, R_N + R_E)$ contains the individual and ethical reward functions:

- R_0 (individual reward): promotes collecting as many valid answers as possible.

$$R_0(s,a,s') \doteq \begin{cases} 1, & \text{if } a{=}Ask \text{ and } valid\_answer(s') \\ 0, & \text{otherwise} \end{cases}$$

- R_N (normative reward): punishes i) asking questions when the user is engaged or provides non-valid or slow answers; and ii) waiting (i.e., not asking questions) when the user is not engaged, as these moments of low engagement should not be wasted:

$$R_N(s,a,s') \doteq \begin{cases} -2, & \text{if } ((a{=}Ask \text{ and } (engaged(s) \text{ or not } valid\_answer(s') \text{ or } slow\_answer(s'))) \\ & \quad \text{or } (a{=}Wait \text{ and not } engaged(s))) \\ 0, & \text{otherwise} \end{cases}$$

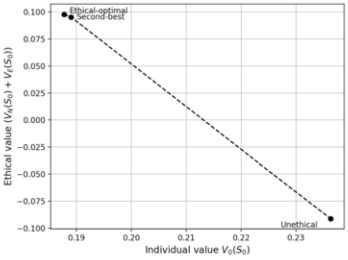
- R_E (evaluative reward): promotes asking questions that get a quick and valid response without interrupting engagement:

$$R_E(s,a,s') \doteq \begin{cases} 1, & \text{if } (a{=}Ask \text{ and } quick\_answer(s') \text{ and } valid\_answer(s') \text{ and not } engaged(s)) \\ 0, & \text{otherwise} \end{cases}$$

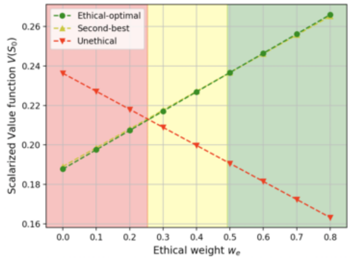
Thus $R_N + R_E$ encapsulates our notion of respect applied to the context of performing in-game questionnaires. Finally, state transition probabilities in $T(s,a,s')$ are approximated by observing the frequencies of such transitions in 500 game executions.

Next, we apply the ethical embedding algorithm. Figure 4a visualizes the convex hull, that is, those policies that are maximal for some value of w_e. Specifically, black dots signal the ethical-optimal policy ($\vec{V}^*$, the one that maximizes the ethical value function

[3]Notice that in our simple Pong game, engagement can be assumed if the user moves the paddle, but this varies for different games. Moreover, although moving the paddle can only be done in-game, and thus we assume low engagement in menus, it may also happen if the play is slow enough.



(a) The convex hull for our ethical MOMDP.



(b) Scalarized policies in the weight space.

Figure 4. The ethical embedding process: (a) visualizing the convex hull, and (b) finding the ethical weight.

$(V_N + V_E)$); the second-best ethical optimal policy ($\vec{V'}^*$); as well as the (unethical) policy that maximizes the individual value (V_0). Next, we solve Eq. 3 and obtain a value of $w_e > 0.49237$. In fact, this value can be empirically found by plotting, as in Figure 4b, the scalarised values for these tree policies, and by identifying the value of w_e for which the ethical-optimal policy has the highest scalarised value (and this is also the case for all w_e values in the green area). Then, we set the weight to $w_e = 0.5$ and return the ethical MDP $\langle S, A, R_0 + w_e(R_N + R_E), T\rangle$ as the environment that guarantees that the agent will learn to behave ethically. Finally, it is worth mentioning that Theorem 1 in [19] formally guarantees that the agent will still learn the same ethical optimal policy regardless of the scale[4] of the ethical rewards considered before scalarisation.

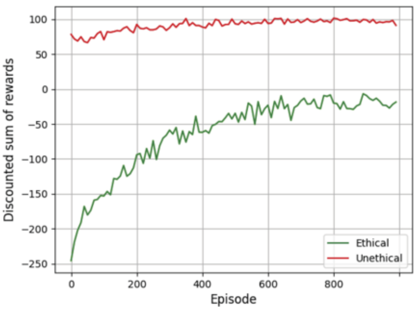
5. Results

The resulting ethical MDP provides a simple environment for our conversational agent to learn to be respectful while asking survey questions. Here, we empirically prove so by applying Q-learning [26]. Specifically, we set a learning rate $\alpha = 0.7$, a discount factor $\gamma = 0.7$, and an ε-greedy policy for exploration along 1000 episodes, where each episode corresponds to a playthrough of our three-level Pong game[5].

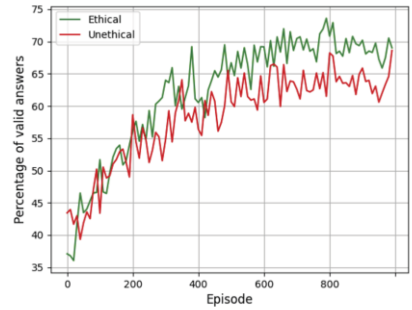
To better assess the impact of the ethical embedding. Figure 5a illustrates the convergence, in terms of the accumulated reward, of the learning of two agents: in green, our ethical agent; in red, an unethical agent that just considers the individual reward R_0. Not surprisingly, our ethical agent takes longer to learn, and accumulates negative rewards as the R_N reward is quite demanding and punishes the agent for not taking advantage of all low engagement situations in slow play. However, this does not preclude our ethical agent to elicit necessary information. In fact, as depicted in Figure 5b, once it learns, it manages to get more valid answers than the unethical agent, which relies on the user to answer questions even if interrupted.

Beyond checking that the ethical agent manages to accomplish its individual objective, we need to assess it learns a respectful behaviour, asking questions when the user's engagement is low, which typically happens while the user is in menus. Thus, we focus on comparing the number of questions prompted in-game and in menus. Specifically, Figure 5c shows how the green ethical agent manages to drastically reduce the number of questions in-game (as opposed to the red unethical agent) and Figure 5d shows how the

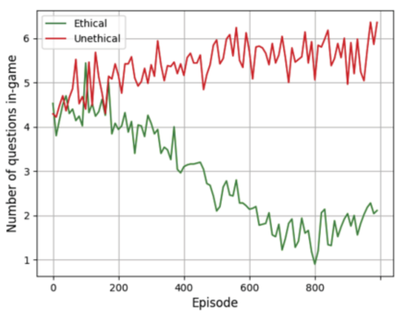
[4] As long as the reward of praiseworthy actions are > 0 and the ones for blameworthy actions are < 0.
[5] Our code is publicly available at https://github.com/ericRosello/EthicalCA.



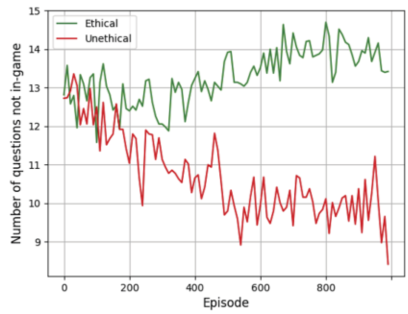
(a) Accumulated discounted reward.



(b) % of questions that received a valid answer.



(c) Number of questions in-game.



(d) Number of questions in menus.

Figure 5. Evolution of different metrics throughtout the learning process.

ethical agent focuses in asking most of the questions in menus (a behaviour that again contrasts with the one of the unethical agent). Thus, overall, we can claim that our conversational agent has successfully learnt to ask survey questions without disturbing the user play, that is, behaving in alignment with the moral value of respect.

6. Conclusions and Future Work

This paper proposes an ethical conversational agent in charge of gathering User eXperience data while the user is playing a game. The agent, applying the ethical embeddding method, learns to respectfully conduct the in-game questionnaire. This method transforms an ethical MOMDP into an ethical MDP that can be addressed by standard RL algorithms. Specifically, we defined the learning environment based on the Pong game, and used Q-learning with a simulated user to assess the ethical agent's learning. The results show that our ethical agent asks the user questions in more appropriate situations (low user engagement) than the unethical agent. Thus, it fulfils the ethical objective while still pursuing the individual one (i.e obtain as much UX data as possible). Indeed, the ethical agent obtained a higher proportion of valid answers than the unethical one, while reducing gameplay interruptions.

Future work should explore the generalization of our approach to alternative games and virtual reality experiences, as the activity of the user (and so engagement) is highly dependent on the (game) mechanics. The study of other moral values (e.g. fairness) is another interesting line of research.

References

[1] D. Alexandrovsky, S. Putze, M. Bonfert, S. Höffner, P. Michelmann, D. Wenig, R. Malaka, and J. D. Smeddinck. Examining design choices of questionnaires in vr user studies. In *CHI'20, 1–21*, 2020.

[2] M. Anderson, S. L. Anderson, and C. Armen. An approach to computing ethics. *IEEE Intelligent Systems*, 21(4):56–63, 2006.

[3] T. Arnold, D. Kasenberg, and M. Scheutz. Value alignment or misalignment - what will keep systems accountable? In *AAAI Workshops*, 2017.

[4] L. Barrett and S. Narayanan. Learning all optimal policies with multiple criteria. *Proc. of 25th ICML*, pages 41–47, 01 2008.

[5] A. Beck, B. Stevens, K. A Bard, and L. Cañamero. Emotional body language displayed by artificial agents. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2(1):1–29, 2012.

[6] A. Bignold, F. Cruz, R. Dazeley, P. Vamplew, and C. Foale. An evaluation methodology for interactive reinforcement learning with simulated users. *Biomimetics*, 6(1):13, 2021.

[7] J. Casas-Roma and J. Conesa. Towards the design of ethically-aware pedagogical conversational agents. In *Int. Conference on 3PGCIC*, pages 188–198. Springer, 2020.

[8] J. Cassell. Embodied conversational agents: representation and intelligence in user interfaces. *AI magazine*, 22(4):67–67, 2001.

[9] A. Cheng and K. R. Fleischmann. Developing a meta-inventory of human values. *Proc. of the ASIS&T*, 47(1):1–10, 2010.

[10] X. Han, M. Zhou, M. J. Turner, and T. Yeh. Designing effective interview chatbots: Automatic chatbot profiling and design suggestion generation for chatbot debugging. In *Proc. of CHI'21*, pages 1–15, 2021.

[11] M. Hassenzahl. User experience and experience design. *The encyclopedia of HCI*, 2, 2013.

[12] J. R. Keebler, W. J. Shelstad, D. C. Smith, B. S. Chaparro, and M. H. Phan. Validation of the guess-18: a short version of the game user experience satisfaction scale. *J. of Usability Studies*, 16(1):49, 2020.

[13] E. L. Law. The measurability and predictability of user experience. In *ACM SIGCHI EICS, 1–10*, 2011.

[14] E. Levin, R. Pieraccini, and W. Eckert. A stochastic model of human-machine interaction for learning dialog strategies. *IEEE Transactions on speech and audio processing*, 8(1):11–23, 2000.

[15] H. L. O'Brien and E. G. Toms. What is user engagement? a conceptual framework for defining user engagement with technology. *Journal of the ASIS&T*, 59(6):938–955, 2008.

[16] C. Peters, G. Castellano, and S. De Freitas. An exploration of user engagement in hci. In *Proc. of the Int. Workshop on Affective-Aware Virtual Agents and Social Robots*, pages 1–3, 2009.

[17] I. Rodríguez and A. Puig. Open the microphone, please! conversational ux evaluation in virtual reality. In *Workshop 'Evaluating user experiences in mixed reality' in CHI'21*, 2021.

[18] M. Rodriguez-Soto, M. Serramia, M. Lopez-Sanchez, and J. A. Rodriguez-Aguilar. Instilling moral value alignment by means of multi-objective reinforcement learning. *Ethics and Information Technology*, 24(1):1–17, 2022.

[19] Manel Rodriguez-Soto, Maite Lopez-Sanchez, and Juan A Rodriguez-Aguilar. Multi-objective reinforcement learning for designing ethical environments. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence*, pages 1–7, 2021.

[20] D. M. Roijers, P. Vamplew, S. Whiteson, and R. Dazeley. A survey of multi-objective sequential decision-making. *J. Artif. Int. Res.*, 48(1):67–113, October 2013.

[21] J. Schatzmann, K. Weilhammer, M. Stuttle, and S. Young. A survey of statistical user simulation techniques for reinforcement-learning of dialogue management strategies. *The knowledge engineering review*, 21(2):97–126, 2006.

[22] K. Scheffler and S. Young. Automatic learning of dialogue strategy using dialogue simulation and reinforcement learning. In *Proc. of HLT*, volume 2, 2002.

[23] S. H. Schwartz. An overview of the schwartz theory of basic values. *Online readings in Psychology and Culture*, 2(1):2307–0919, 2012.

[24] V. J. Shute. Stealth assessment in computer-based games to support learning. *Computer games and instruction*, 55(2):503–524, 2011.

[25] A. Steinmaurer, M. Sackl, and C. Gütl. Engagement in in-game questionnaires-perspectives from users and experts. In *2021 7th Int. Conference of the iLRN*, pages 1–7. IEEE, 2021.

[26] R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[27] W. Wallach and C. Allen. *Moral machines: teaching robots right from wrong*. Oxford Univ. press, 2008.

[28] H. Yu, Z. Shen, C. Miao, C. Leung, V. R. Lesser, and Q. Yang. Building ethics into artificial intelligence. In *IJCAI*, page 5527–5533, 2018.