ADDRESSING TRUST CONCERNS IN EDUCATIONAL ENVIRONMENTS: DEVELOPING AN EXPLAINABLE EMBODIED CONVERSATIONAL AGENT

N. Sautchuk-Patricio, P. Henning

Hochschule Karlsruhe (GERMANY)

Abstract

Technological advancements have significantly impacted educational technologies, offering new avenues to engage learners effectively. Virtual Reality (VR) holds promise in this regard, providing personal and social affordances crucial for immersive educational experiences.

Interfaces in educational technologies play a pivotal role in enhancing engagement and comprehension. VR interfaces increasingly incorporate avatars representing users and Embodied Conversational Agents (ECAs) simulating human-like interactions through natural language processing and visually appealing representations. For instance, in distance education, the lack of personal and social presence is a noted concern. ECAs can address this by enhancing social interaction and motivation.

While ECAs offer educational potential, trust issues arise. Chiou et al. present a study suggesting that trust impacts learning concepts and retention. Therefore, ensuring trust is also an important factor in the educational process. Explainability methods can mitigate trust concerns by explaining their reasoning (justification), explaining their view of the environment (internal state), or explaining their plans (intent). Explainable ECAs fall under eXplainable Artificial Intelligence, using similar techniques but with the difference of aiming for autonomous interaction with humans and the environment. Verbal and non-verbal modes of explainability in ECAs offer users diverse channels for comprehension, enhancing trust and enriching user experiences. Specifically, verbal explainability methods, frequently utilizing text-to-speech technologies, offer spoken interpretations of complex information within ECAs.

This paper will explore the efficacy of explainability methods in ECAs within educational environments, assessing their impact on trust, user motivation and engagement. We propose a customizable ECA for educational platforms, offering varied personalities and tones of voice —ranging from enthusiastic to apathetic— during the explanation. Users can tailor the ECA's emotional tones and voice characteristics, including pitch, tempo, gender, and frequency, ensuring a responsive interaction experience aligned with instructional preferences and emotional context.

Keywords: Embodied Conversational Agent, Explainability, Trust, Artificial Intelligence.

1 INTRODUCTION

Technological progress has had a significant impact on educational technologies, introducing novel ways to effectively engage learners. Virtual Reality (VR) shows great potential in this regard, offering essential personal and social affordances necessary for an immersive educational experience [1].

Interfaces within educational technologies are crucial for enhancing engagement and comprehension. In the scope of VR, interfaces now integrate avatars to represent users and Embodied Conversational Agents (ECAs) to simulate human-like interactions through natural language processing and visually appealing representations [2]. For example, in distance education, the absence of personal and social presence is a significant concern. ECAs can mitigate this by bolstering social interaction and motivation [3].

Although ECAs present educational promise, concerns regarding trust emerge. Chiou et al. conducted a study indicating that trust influences learning concepts and retention [4]. Hence, ensuring trust is a crucial element in the educational process. Explainability methods can alleviate trust concerns by providing explanations for their reasoning (justification), articulating their perception of the environment (internal state), or elucidating their plans (intent) [5].

Explainable ECAs are categorized within eXplainable Artificial Intelligence, employing similar techniques but with the objective of facilitating autonomous interaction with humans and the environment. Through both verbal and non-verbal modes of explainability, ECAs provide users with varied ways for understanding, thereby bolstering trust and enhancing user experiences. Verbal explainability methods, often leveraging text-to-speech technologies, giving spoken interpretations of intricate information within ECAs.

This study aims to investigate the potential impact of emotional settings of Explainable ECAs on trust, information retention, and motivation levels of participants in educational settings. The research hypotheses for this study are as follows:

H1: Participants are expected to exhibit higher levels of trust in an ECA configured with a "happy setting" compared to one with a "sad setting".

H2: It is hypothesized that participants will retain more information from a lecture delivered by a happy ECA than a sad ECA.

H3: Participants are anticipated to report feeling more motivated when interacting with a happy ECA compared to one with a "sad setting".

2 METHODOLOGY

2.1 Participants and design

The study utilized a between-subjects design and employed a randomized sampling method to ensure representative participation from the target population. A total of 40 participants were randomly selected from a pool of potential candidates. Participant recruitment was facilitated through Prolific [6], an online platform that connects researchers with a diverse pool of participants.

The first group consisted of 20 individuals who interacted with an ECA within a sad context, while the second group of 20 individuals engaged with a happy ECA. Out of the total participants, 24 are female and 16 are male, evenly distributed among the different scenarios. The average age of the participants was 27.9 years. In terms of education level, 35% hold a university degree, 25% are currently enrolled in university, and 12.5% possess a Master's degree. The participants are from various countries, including Canada, Denmark, France, Latvia, Mexico, New Zealand, Poland, Portugal, Romania, South Africa, the United Arab Emirates, and the United Kingdom. Among the total participants, 32.5% speak English as their first language, 20% speak Portuguese, 17.5% speak Polish, and 15% speak Spanish.

2.2 Materials

The materials included a pre-questionnaire, followed by a learning activity involving interaction with an ECA, a retention test to assess comprehension, and finally a post-questionnaire to evaluate participant perceptions.

2.2.1 Pre-questionnaire

Before engaging in the learning activity, participants completed a pre-questionnaire consisting of sociodemographic questions, including gender, age, nationality, mother tongue, level of education, as well as their current emotional status based on a question with six options related to the basic emotions [7], and previous knowledge related to the topic of the study.

2.2.2 Learning activity

Participants are presented to a learning activity with a video featuring an ECA delivering a lecture on a topic related to Brazilian History. The lecture was delivered in English and lasted approximately 3 minutes. The ECA featured a 3D character and animations from Mixamo [8], developed using Unity. The voices were generated using Murf AI [9] based on the same written script.

As shown in Figure 1, both scenarios featured the same male ECA within an identical background, employing the same sequence of animations. The sole distinction between these scenarios was the voices used, with one expressing sadness and the other demonstrating happiness.

Figure 1. Learning activity with ECA

2.2.3 Retention test

Following the learning activity, participants undergo a retention test comprising eight multiple-choice questions with 5 alternatives based on the content of the lecture. This test assesses participants' comprehension and retention of the material presented by ECA. The same questions were asked in the two scenarios.

2.2.4 Post-questionnaire

After completing the retention test, participants respond to a post-questionnaire. Initially, participants indicate their current emotional state, choosing from the six basic emotions as used in the pre-questionnaire. Drawing from the Cognitive Affective Model of Learning with Instructional Video, subsequent statements assess whether the learner identifies the instructor's emotion, responds to it, and ultimately assimilates the instructor's emotional state [10]. In assessing trust, participants utilized a questionnaire adapted from [11].

Participants also assessed the ECA based on four factors outlined in the Agent Persona Instrument (API): Facilitating Learning (10 items), Credible (5 items), Human-like (5 items), and Engaging (5 items) [12]. The API aims to measure the ECA's effectiveness as an instructor, emphasizing its role in facilitating learning and bolstering credibility to enhance learners' cognitive processes, reflection, and comprehension. Additionally, the human-like factor evaluates the ECA's natural communication abilities, encompassing emotional expression and nonverbal cues, while the engagement factor evaluates the agent's positive social presence during interaction with the learner.

All items of the post-questionnaire were presented with a 5-point Likert scale, ranging from 1=Strongly disagree to 5=Strongly agree.

3 RESULTS

Initially, Cronbach's Alpha was computed to assess the internal consistency of the utilized scales. All Cronbach's Alpha were good or excellent (Facilitating Learning: 0.95; Credible: 0.85; Human-Like: 0.93; Engaging: 0.95; Motivation: 0.83; Trust: 0.92).

In Figures 2 and 3, graphical summaries of the results for the various measured factors are presented, categorized by gender. Upon analyzing interactions in the two scenarios separately by gender, distinct patterns emerge for each gender. Female participants tend to perceive the sad ECA as more credible, trusted and human-like, while male participants view it as such in the happy scenario. Regarding motivation, women express higher motivation with the happy ECA, whereas men are more motivated in the sad scenario. Both female and male participants agree that the sad ECA would facilitate learning, while the happy ECA is deemed more engaging.

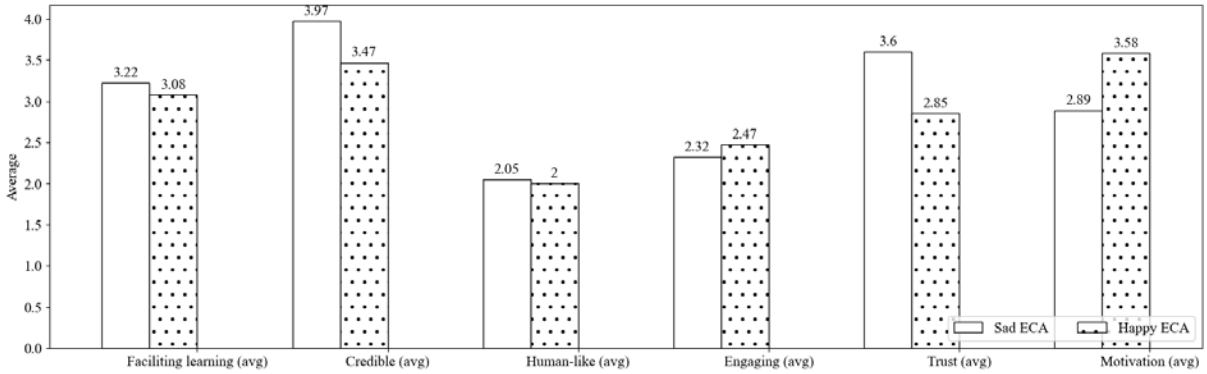

Figure 2. Female participants result with ECA

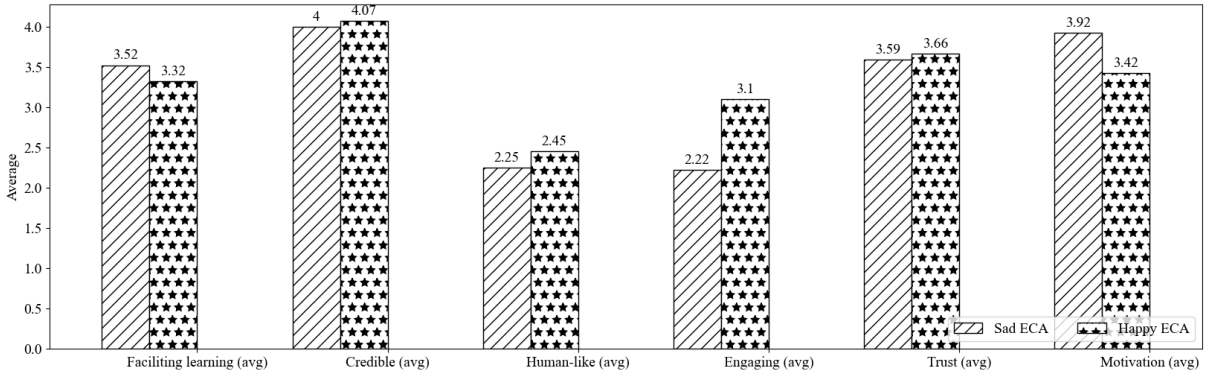

Figure 3. Male participants result with ECA

Taking into account the score of the retention test applied after the learning activity, it is possible to see a trend in Table 1 towards better results in the case of happy ECA. Women were better in both scenarios, but the trend holds for both genders.

Table 1. Retention score average

	Sad ECA		***Happy ECA***	
	Female	***Male***	***Female***	***Male***
Retention score average (among 0 and 1)	0.79	0.76	0.85	0.81

A correlation analysis was conducted among various variables measured in the study. Based on Spearman's rank correlation coefficient (rho), moderate correlation is indicated when rho ranges between 0.4 and 0.59, strong correlation is observed when the value falls between 0.6 and 0.79, and very strong correlation is identified when the coefficient exceeds 0.8. Tables 2 and 3 display the outcomes of the correlation analysis among the tested factors for both scenarios, highlighting those with at least moderate correlation.

When contrasting the happy scenario with the sad one, it becomes apparent that trust exhibits a strong correlation with all factors. However, in the sad scenario, trust is only strongly correlated with credibility, whereas in the happy setting, it demonstrates strong correlations with facilitating learning, credibility, human-likeness, and engagement. Moreover, it is possible to note in both tables that there is strong correlation between different factors.

Table 2. Correlation analysis based on Spearman's rank correlation coefficient- Happy ECA

	Facilitating Learning	***Credible***	***Human-like***	***Engaging***	***Trust***
Motivation	Moderate (0.59)	Strong (0.66)	-	-	Moderate (0.54)
Facilitating Learning	-	Strong (0.65)	Very strong (0.8)	Strong (0.75)	Strong (0.78)
Credible	-	-	Strong (0.7)	Strong (0.6)	Very strong (0.82)
Human-like	-	-	-	Very strong (0.91)	Very strong (0.8)
Engaging	-	-	-	-	Strong (0.73)

Table 3. Correlation analysis based on Spearman's rank correlation coefficient - Sad ECA

	Facilitating Learning	***Credible***	***Human-like***	***Engaging***	***Trust***
Motivation	Strong (0.77)	Strong (0.67)	-	-	-
Facilitating Learning	-	Very strong (0.91)	Weak (0.36)	Moderate (0.41)	Moderate (0.59)
Credible	-	-	Moderate (0.41)	Moderate (0.46)	Strong (0.67)
Human-like	-	-	-	Very strong (0.88)	Moderate (0.59)
Engaging	-	-	-	-	Moderate (0.59)

A descriptive analysis was conducted to examine the relationship between various factors and the retention test score. Figures 5 to 9 illustrate that in the happy ECA scenario, there is a positive correlation between the measured factors and the retention test score, whereas in the sad ECA scenario, the correlation is negative. Figure 4 indicates a positive correlation between motivation and the retention test score in the happy ECA scenario, while in the sad ECA scenario, there is no correlation between these two factors.

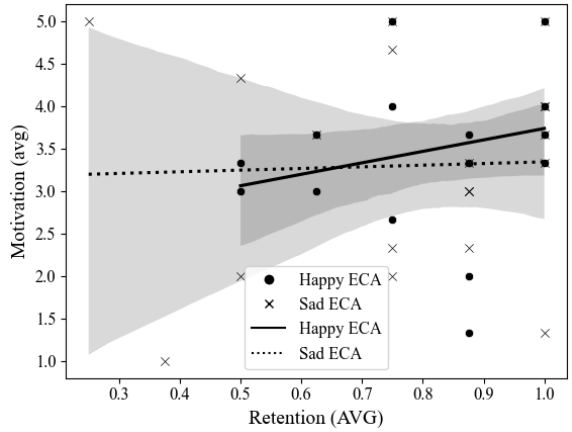

Figure 4. Correlation among Motivation and Retention

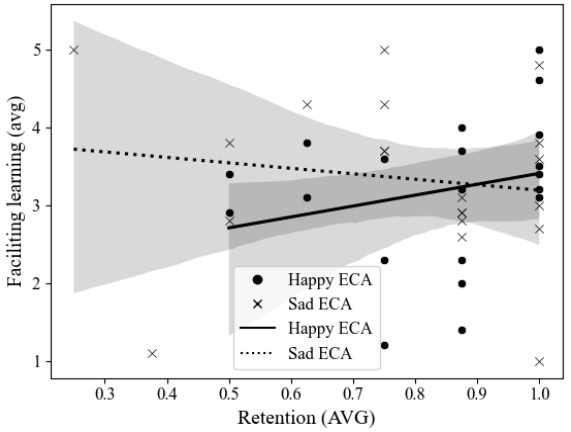

Figure 5. Correlation among Facilitation of Learning and Retention

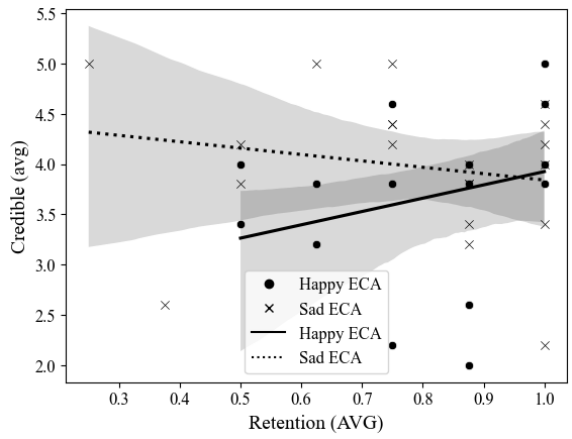

Figure 6. Correlation among Credibility and Retention

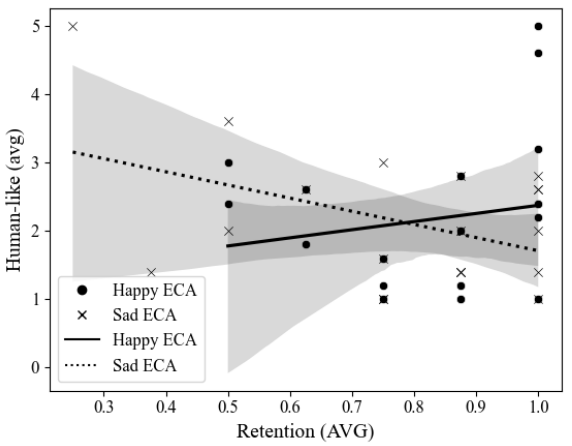

Figure 7. Correlation among Human-likeness and Retention

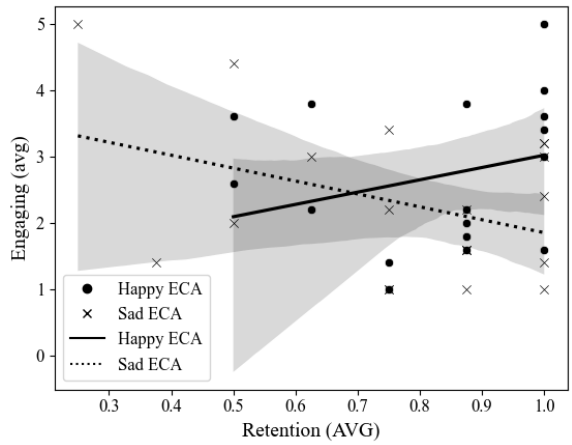

Figure 8. Correlation among Engagement and Retention

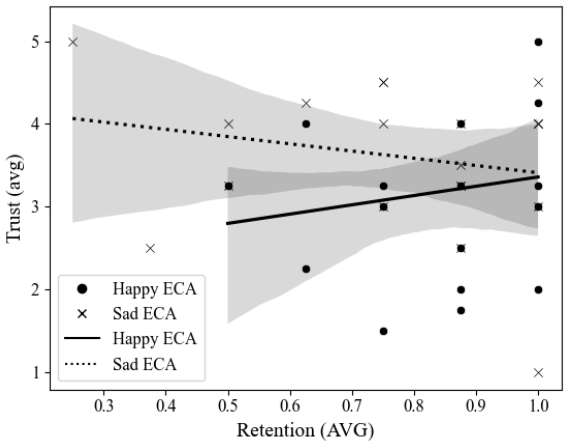

Figure 9. Correlation among Trust and Retention

4 CONCLUSIONS

In this exploratory study, we aimed to investigate the impact of an Explainable ECA configured with either a "happy setting" or a "sad setting" on participants' trust, retention of information, and motivation. While a previous study has explored the correlation between the voice of an ECA and trust as well as learning outcomes, it primarily assessed the voice quality [4], unlike the current study, which specifically examined the emotion conveyed by the voice. While our findings provide valuable insights, several limitations should be acknowledged.

One notable limitation is the small number of participants per scenario, which may have affected the robustness of our results. Additionally, the interaction with the ECAs was conducted via video, which limited the ability for participants to "really" interact with the ECAs as they would with a real instructor. Moreover, the retention test administered immediately after the lecture may not accurately reflect long-term retention of information.

Our hypotheses yielded mixed results. While we anticipated that participants would exhibit higher levels of trust in the ECA with a "happy setting," this was only true for male participants. Female participants, on the other hand, showed higher levels of trust in the sad ECA scenario. Upon examining the data, it became apparent that men exhibited higher levels of trust compared to women in both scenarios. Similarly, while we expected participants to retain more information from the lecture delivered by the ECA with a "happy setting," this was indeed the case for both genders, albeit with a modest difference in scores between the two scenarios. Regarding motivation, our hypothesis held true for male participants, but female participants reported higher levels of motivation in the sad ECA scenario. Analyzing the data, it was evident that men displayed higher levels of motivation in both scenarios

compared to their female counterparts. This trend could potentially be attributed to the fact that the ECA used in the study was male, a pattern observed in [13].

For future work, it would be valuable to conduct tests using ECAs of both genders to explore potential gender-based differences in participant feelings. Additionally, altering parameters beyond just the voice, such as gestures and facial expressions, could provide a more nuanced expression of sadness and happiness. Improving the human-likeness of the ECAs, particularly in terms of facial expressions, may also enhance participants' engagement and interaction with the agents. Overall, further research is needed to fully understand the complex interplay between Explainable ECAs' emotional settings and trust concerns in their use in educational environments.

REFERENCES

[1] F. Grivokostopoulou, K. Kovas, and I. Perikos, “The Effectiveness of Embodied Pedagogical Agents and Their Impact on Students Learning in Virtual Worlds,” *Applied Sciences*, vol. 10, no. 5, 2020. Retrieved from https://doi.org/10.3390/app10051739.

[2] H. M. Aljaroodi, M. T. P. Adam, R. Chiong, and T. Teubner, “Avatars and Embodied Agents in Experimental Information Systems Research: A Systematic Review and Conceptual Framework,” *Australasian Journal of Information Systems*, vol. 23, 2019. Retrieved from https://doi.org/10.3127/ajis.v23i0184.1.

[3] I. S. Fitton, D. J. Finnegan, and M. J. Proulx, “Immersive virtual environments and embodied agents for e-learning applications,” *PeerJ Computer Science*, vol. 6, Nov. 2020. Retrieved from https://doi.org/10.7717/peerj-cs.315.

[4] E. K. Chiou, N. L. Schroeder, S. D. Craig. “How we trust, perceive, and learn from virtual humans: The influence of voice quality,” Computers & Education, vol. 146, 2020. Retrieved from https://doi.org/10.1016/j.compedu.2019.103756.

[5] S. Wallkötter, S. Tulli, G. Castellano, A. Paiva, and M. Chetouani, “Explainable Embodied Agents Through Social Cues: A Review,” *ACM Transactions on Human-Robot Interaction,* vol. 10, pp 1–24, 2021. Retrieved from https://doi.org/10.1145/3457188.

[6] Prolific, Accessed 08 May 2024. Retrieved from https://www.prolific.com/.

[7] P. Ekman. (1992). “An Argument for Basic Emotions,” *Cognition and Emotion*, 6(3/4), pp.169-200, 1992. Retrieved from https://doi.org/10.1080/02699939208411068.

[8] Mixamo, Accessed 08 May 2024. Retrieved from https://www.mixamo.com/.

[9] Murf AI, Accessed 08 May 2024. Retrieved from https://murf.ai/.

[10] T. Horovit, R. E. Mayer. “Learning with human and virtual instructors who display happy or bored emotions in video lectures,” *Computers in Human Behavior*, vol. 119, 2021. Retrieved from https://doi.org/10.1016/j.chb.2021.106724.

[11] C.B. Nordheim, A. Følstad, C.A. Bjørkli. “An Initial Model of Trust in Chatbots for Customer Service—Findings from a Questionnaire Study,” *Interacting with Computers*, vol. 31, no. 3, pp. 317–335, 2019. Retrieved from https://doi.org/10.1093/iwc/iwz022.

[12] J. Ryu, A. L. Baylor. “The Psychometric Structure of Pedagogical Agent Persona,” *Technology, Instruction, Cognition and Learning*, vol. 2, pp. 291-314, 2005.

[13] G. Ozogul, A.M. Johnson, R.K. Atkinson, M. Reisslein. “Investigating the impact of pedagogical agent gender matching and learner choice on learning outcomes and perceptions,” *Computers & Education*, vol. 67, No. C, pp 36–50, 2013. Retrieved from https://dl.acm.org/doi/10.5555/2753875.2753958.