

ARTICLE OPEN

Trust and acceptance of a virtual psychiatric interview between embodied conversational agents and outpatients

Pierre Philip^{1,2,3*}, Lucile Dupuy^{1,3}, Marc Auriacombe^{1,3,4}, Fushia Serre^{1,3,4}, Etienne de Sevin^{1,3}, Alain Sauteraud^{2,3} and Jean-Arthur Micoulaud-Franchi^{1,2,3}

Virtual agents have demonstrated their ability to conduct clinical interviews. However, the factors influencing patients' engagement with these agents have not yet been assessed. The objective of this study is to assess in outpatients the trust and acceptance of virtual agents performing medical interviews and to explore their influence on outpatients' engagement. In all, 318 outpatients were enrolled. The agent was perceived as trustworthy and well accepted by the patients, confirming the good engagement of patients in the interaction. Older and less-educated patients accepted the virtual medical agent (VMA) more than younger and well-educated ones. Credibility of the agent appeared to main dimension, enabling engaged and non-engaged outpatients to be classified. Our results show a high rate of engagement with the virtual agent that was mainly related to high trust and acceptance of the agent. These results open new paths for the future use of VMAs in medicine.

npj Digital Medicine (2020)3:2; <https://doi.org/10.1038/s41746-019-0213-y>

INTRODUCTION

Mental disorders are chronic conditions requiring repetitive and time-consuming psychiatric consultations. Because of their increasing prevalence in Western societies and the shortage of physicians, there is a need for innovative clinical solutions to interview patients without mobilising human resources.¹ Digital medicine is a promising solution to conduct patients' follow-up, and virtual medical agents (VMAs) have been considered as potential candidates to assist real physicians.^{2–4} VMAs are embodied conversational agents (ECAs), defined in the human–computer interface (HCI) field as computer characters able to engage in face-to-face dialogue through verbal and nonverbal behaviour.⁵ ECAs have already been used in medicine for the diagnosis,^{6–8} prevention,^{9,10} and treatment of medical conditions.^{11,12} Taken together, the above studies tend to show that ECAs have a positive effect, as they rely on a virtual face-to-face interaction, thus creating empathy for users and facilitating disclosure of negatively connoted symptoms such as addiction, depression, and post-traumatic stress disorder manifestations.^{6–8} In addition, ECAs can significantly help healthcare delivery by saving physicians' time, diminishing the variability between professional interventions, providing 24/7 and easy-to-access support and monitoring of patients, and by managing huge volumes of medical data.

Nevertheless, to achieve these medical objectives consistently over the long term, i.e., management of chronic diseases, researchers need to support patients' engagement with VMAs¹³ by fostering a “strong doctor-patient relationship”.¹⁴ In a recent publication by the Lancet psychiatry commission,¹⁴ experts underlined two major dimensions that will consolidate engagement with health technologies: acceptance and trust. Acceptance and trust are complex and interrelated constructs evaluated with various tools in the HCI literature, and therefore need a clear theoretical background. In their article, Bhugra and colleagues suggest that acceptance can be promoted by providing easy ways to “entry and retrieval of data”, and by being “enjoyable to use”. In

the HCI literature, these constructs are commonly referred to as the concepts of perceived ease of use (also called by some author usability¹⁵) and perceived usefulness (or satisfaction),^{15–17} which altogether can predict the use or rejection of technology in general^{15,16} and ECAs in particular.^{18,19}

The second issue raised by Bhugra and colleagues is trust in the digital device, so that “patients feel confident in sharing their psychiatric history”,¹⁴ which is also a well-studied construct and a central aspect of engagement with technologies. In the HCI literature, Cassell & Bickmore²⁰ proposed that trust in an ECA occurs if the user perceives credibility and benevolence in the agent. Credibility corresponds to the belief that the agent has the ability and the expertise in a specific domain to perform the task, and benevolence means that the agent cares about and will act according to one's interest. In terms of ECA design, Cassell & Bickmore suggest that credibility can be shown to the user by using expert vocabulary, adapted appearance, or professional affiliation, whereas benevolence can be demonstrated through greetings, small talk, or references to past experiences. In a non-medical context, some studies have shown that perceived credibility and benevolence can predict acceptance of wearable technologies²¹ and ECA usage for e-commerce.^{22,23} Thus, acceptance and trust are central in the evaluation of new technologies and should be assessed systematically in the context of ECAs for medical care. Surprisingly, there is a lack of standardised and validated scales to rigorously measure these dimensions, in particular in medicine.¹³

Furthermore, engagement, acceptance, and trust do not depend only on the design of the software but may vary highly depending on the characteristics of the user. Age, gender, education, and health conditions of users can influence engagement, acceptance, and trust in technologies but have been insufficiently studied.^{24,25} In addition, the impact of the medical domain covered by the agent (i.e., type of disease targeted) on engagement, acceptance, and trust in a similar VMA has never been studied. Therefore, the characteristics of the user and the

¹University of Bordeaux, Bordeaux, France. ²Clinique du Sommeil, CHU de Bordeaux, Bordeaux, France. ³SANPSY, CNRS USR 3413, Bordeaux, France. ⁴Pôle addictologie, CH Charles Perrens and Unité de Soins Complexes d'addictologie (USCA) CHU de Bordeaux, Bordeaux, France. *email: pierre.philip@u-bordeaux.fr

medical domain covered by the agent need to be considered when evaluating a particular technology.

Recently, our team published several articles demonstrating that ECAs can conduct reliable and valuable clinical interviews and make psychiatric diagnoses (depression and addiction) in outpatients seen in a sleep clinic.^{6,7,26} In addition, we showed that a VMA was better accepted than a questionnaire displayed on a tablet to diagnose major depressive disorder (MDD).²⁵ In this new study, we explored the impact of both the characteristics of the user and the context of the psychiatric interview covered by the VMA (for depression or addiction screening) on engagement, acceptance, and trust, and attempted to determine the threshold of acceptance and trust in VMAs that are associated with positive engagement.

RESULTS

Sample characteristics

A total sample of 318 patients were analysed. Patients' characteristics are summarised in Table 1.

Group comparisons showed no significant differences between patients included in Study 1 and Study 2 regarding demographic or sleep disorder characteristics (all $p > 0.100$). We therefore grouped the two populations to analyse characteristics.

Globally, participants were middle-aged (45.01 years on average; SD = 13.33), with one third aged over 50 years old. There was a slightly higher educational level than the French population average,²⁷ with about half of the participants having a bachelor's degree. Most participants suffered from nocturnal breathing disorders, which matches the prevalence among the general population.²⁸ Hypersomnia and narcolepsy were also well represented, with ~10% of our sample suffering from these disorders. Finally, about a third of participants suffered from non-organic sleep complaints, such as asymptomatic snoring, transient insomniac sleep complaints, or sleep hygiene disorders.

Trust, acceptance, and engagement with the VMA

Acceptance and trust data are presented in Fig. 1. Acceptance of the overall system (score of the AES scale) was rated very

positively, with 68.2% of patients being "very satisfied" by the VMA's usability, and 78.1% of patients rating the VMA more than three out of five for satisfaction. Regarding trust (score on the ETQ scale), the VMA was perceived as trustworthy to perform medical interviews. Indeed, 68.2% of patients "totally agreed" that the VMA was benevolent, and 79.2% of patients rated the VMA more than two out of three for credibility.

The distribution of patients' answers to the Engagement question is presented in Fig. 2. More than half (57.23%) of outpatients were willing to interact with the VMA in the future, suggesting a positive engagement with the agent.

Table 1. Sample characteristics.

	Total (N = 318)
Age (M (SD))	45.01 (13.33)
<30 years (%)	19.9%
30–50 years (%)	38.8%
>50 years (%)	42.0%
Gender (% males)	45%
Education (in years) (M (SD))	13.36 (2.98)
Middle school (%)	10.3%
High school (%)	38.6%
University degree (%)	50.9%
Type of sleep disorder (%)	
Nocturnal breathing disorders	42.3%
Narcolepsy, hypersomnia	10.7%
Insufficient sleep syndrome	1.9%
Periodic leg movements and RLS	1.3%
Insomnia, ADHD, parasomnia	9.5%
Non-organic sleep complaints	34.4%

RLS restless legs syndrome, ADHD attention deficit hyperactivity disorder

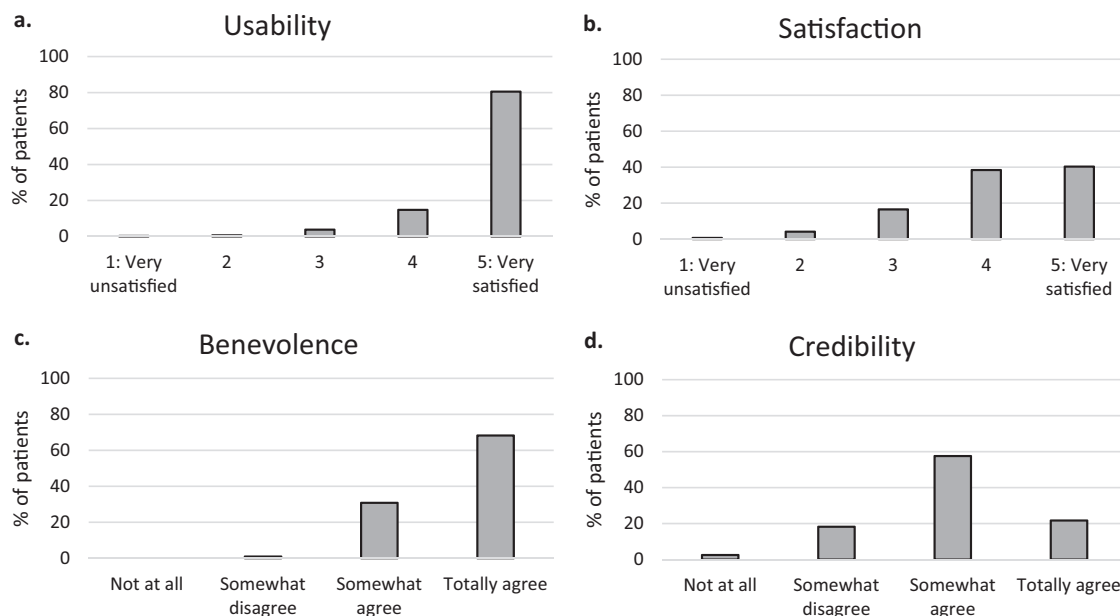


Fig. 1 Distribution of usability, satisfaction, benevolence, and credibility perception. **a** percentage of patients' rating for usability dimension (AES sub-score), **b** percentage of patients' rating for satisfaction dimension (AES sub-score), **c** percentage of patients' rating for benevolence dimension (ETQ sub-score), **d** percentage of patients' rating for credibility dimension (ETQ sub-score).

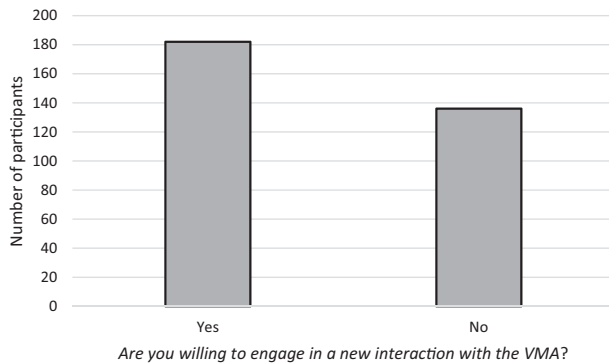


Fig. 2 Distribution of outpatient participants regarding their perceived engagement with the VMA.

Influence of patients' characteristics and VMA interview on acceptance, trust, and engagement

Regarding usability sub-scores, Pearson correlation analyses revealed that age was significantly correlated with the usability sub-score of the AES, with older participants perceiving the system as easier to use than younger ones ($r = 0.143$; $p = 0.11$). Mean comparisons showed that the medical domain covered by the VMA influences usability ($t(316) = -3.385$; $p = 0.001$), and that the addiction interview is perceived as easier to use than interview screening for MDD. Other variables (education, gender, sleep disorder) remained non-significant (all $p > 0.200$), indicating that the system was perceived as easy to use irrespective of these user characteristics. Multivariate analyses (Supplementary Table 1) confirmed the significant influence of age and medical domain ($F(2,316) = 8.587$; $p < 0.001$).

The satisfaction sub-score was significantly correlated with age ($r = 0.169$; $p = 0.011$) and educational level ($r = -0.179$; $p = 0.001$), older and less-educated participants being more satisfied by the system than those younger and with a high level of education. However, users were satisfied by the system regardless of their gender, sleep disorder or medical domain covered by the VMA (all $p > 0.500$). Multivariate analyses (Supplementary Table 1) confirmed the significant influence of age and education ($F(2, 313) = 8.915$; $p < 0.001$).

Concerning trust dimensions, neither credibility sub-score nor benevolence sub-scores seemed to be influenced by any user characteristics or type of psychiatric interview conducted by the VMA. Therefore, we did not perform any multivariate analyses.

Finally, regarding engagement question ("willing to engage in a new interaction"), mean and distribution comparisons analyses showed that patients' engagement varied with regard to VMA interview (MDD or addiction) ($\chi^2(1) = 10.156$; $p = 0.001$) and educational level ($t(313) = -1.993$; $p = 0.005$). The interview for depression screening, and a lower level of education were significant predictors of non-engagement with the ECA. Gender, type of sleep disorder, and age were not significantly different between engaged and non-engaged patients. Logistic regression (Supplementary Table 1) confirmed the significant influence of medical domain and education (omnibus chi-square = 15.689, $df = 2$, $p < 0.001$).

Acceptance and trust thresholds associated with future engagement with the VMA

Statistical receiver operating characteristics (ROC) analyses with the four sub-scores of acceptance and trust (usability, satisfaction, benevolence, and credibility) are presented in Fig. 3 and Supplementary Table 2.

Globally, the four scores classified patients efficiently depending on their future engagement, all area under the curve (AUC) being

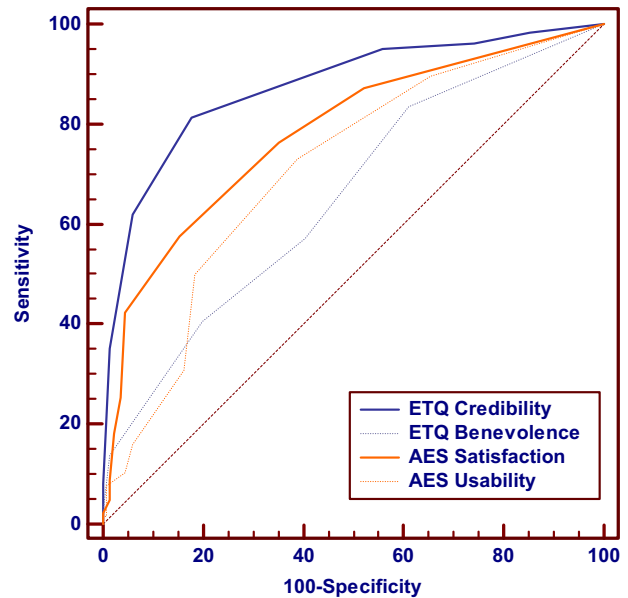


Fig. 3 ROC curves of the four dimensions of trust and acceptance: benevolence, credibility, usability, and satisfaction, to classify engaged and non-engaged subjects.

above 0.650. Analyses revealed cutoff scores of 13.5 and 12.5 for usability and satisfaction, respectively, and 5.5 and 8.5 for credibility and benevolence, in order to classify patients engaged with the VMA. The credibility sub-score of the ETQ obtained significantly the best classification performance compared with the other dimensions (all $p < 0.100$) and showed an AUC of 0.875 ($p > 0.001$), a cutoff sensitivity of 82.4% and a specificity of 81.3%.

DISCUSSION

The objective of this study was to measure engagement, and perceived acceptance and trust of VMAs used to diagnose depression and addiction disorders among outpatients. We explored the influence of outpatients' characteristics and type of medical field covered by the VMA in the variation of engagement, acceptance, and trust. These findings are very encouraging and suggest that VMAs could help a wide range of patients. It also advocates for the consideration of user characteristics and topic of the medical intervention when designing and evaluating ECAs usage. At last, we identified acceptance and trust thresholds to quantify future engagement with the agent.

Taken together, our results show satisfactory levels of acceptance and trust independently of the gender, type of sleep disorders affecting the outpatients, and medical domain covered by the VMA. Previous studies including gender analyses of technology acceptance found contradictory results, some highlighting differences between men and women in acceptance,¹⁹ whereas others did not.²¹ Further studies are needed to investigate the impact of gender of the patient with regard to the gender of the VMA. There was no significant influence of user characteristics and medical domain covered by the VMA on perceived trust, but we observed significant relationships between age, level of education, and perceived acceptance of the system. Contrary to common stereotypes that older adults are less likely to adopt technologies,²⁹ older patients demonstrated greater acceptance of the system displaying the medical agent. Here again, internet technologies are now widely deployed among aging populations to offset the problem of medical deserts and the loss of autonomy of seniors.³⁰ In addition, several older adults

felt that our VMA was like a companion who could help them to manage their health in their remotely situated houses. The public health implications are important, given the aging of the population worldwide and the growing percentage of older individuals with chronic diseases.³¹ Subjects with a low level of education were more satisfied by our system more than those with a higher level. The massive development of internet technologies in all classes of society combined with the widespread availability of “free apps” in the healthcare domain could explain these results.

Furthermore, most of our patients were willing to interact with the agent in the future, suggesting positive engagement. However, we observed that the interview for depression screening and a lower level of education seemed to favour non-engagement with the VMA. Reasons for these influences might be that the conduct of the interview varied between addiction and MDD screening, the former was based on short structured questionnaires (i.e., the CDS-5 and the CAGE questionnaire), whereas the latter involves less-structured questions (i.e., based on the DSM-5 criteria). In line with this hypothesis, our results show that the MDD interview with the VMA was perceived as less easy to use than the addiction interview. Further studies should investigate the impact of length and design of the interview on patients' engagement. In addition, particular attention should be paid to the interaction scenario implemented by the VMA.

Results of the ROC analyses revealed threshold scores for the AES and ETQ scores and sub-scores to detect future engaged and non-engaged users in a clinical setting. This contribution could be helpful for the earlier detection of disengagement with a VMA, especially in a long-term autonomous use (e.g., at home for the follow-up of chronic patients). In particular, we show that the agent's credibility appropriately classified > 80% of patients, which suggests that credibility is the most discriminant dimension in terms of patients' engagement. This result confirms that VMAs can be trusted by patients if used in an appropriate clinical context.⁸ It is in line with findings showing the benefits of therapeutic alliance between patient and physicians^{32,33} and with the suggestion that alliance with digital mental health apps is crucial for the future.¹³ Credibility should therefore be a prime consideration when designing VMAs, especially for chronic disease management.

To promote the use of VMAs in clinical settings, medical and HCI experts and regulatory agencies should work together to identify and adhere to standardised vocabulary, methods for the design^{34–36} and evaluation^{4,13,37} of digital solutions for healthcare. In the longer term, this interdisciplinarity would provide the opportunity to develop VMAs fully compliant with standards and legal aspects, as stated by the International Organization for Standardization (ISO) in the standard for Human-centred design for Interactive systems (ISO 9241-210:2019)^{17,38} and by several health regulatory agencies (such as the National Institute for Health and Care Excellence,³⁹ the National Health Service⁴⁰; the Medical Research Council,⁴¹ or the European Medicines Agency⁴²). In the academic field, among other evaluation models proposed,^{43–45} Torous and colleagues^{46,47} recently recommended four steps to be met before including new technological tools can in clinical practice: (1) consider risk and privacy issues, (2) validate efficacy for health, (3) ensure engagement, and (4) establish interoperability. By evaluating acceptance and trust in a VMA, our study conducted in the context of a psychiatric interview meets the third objective of Torous' framework. Further studies are now needed to compare the validity of the AES and ETQ with that of other evaluation tools,^{13,16,48} whereas keeping in mind the necessity for a common and interdisciplinary vocabulary regarding engagement with technologies. At last, additional studies need to confirm the present results over repeated usage of the VMA in the patient's home. Such studies should investigate the impact of intervention duration, frequency and learnability, severity, and evolution of clinical manifestations, the influence of health and technological

representations by the patients and their professional and informal caregivers, and modalities of integration in healthcare systems, in order to promote the involvement of conversational agents for an efficient patient-centred care.

METHODS

This study follows a quantitative experimental design. Data were collected during two protocols published previously.^{6,7} In study 1,⁶ the objective was to validate the efficacy of a VMA performing MDD diagnosis. Study 2⁷ focused on the validity of the VMA to perform screening and diagnosis for tobacco and alcohol use disorders.

Participants

Participants were recruited among outpatients seen at the Sleep Clinic at the University Hospital of Bordeaux (France) from November 2014 to May 2017. The patient population mirrored the characteristics of a patient group consulting in general practice with a common referral related to sleep complaints and a high rate of co-morbidities, including mood disorders or addiction. They were asked to participate in the study during their clinical interview by a sleep specialist. Gender, age, education, and suspected sleep disorders were collected. In all, 179 participants were recruited for Study 1 (VMA for MDD diagnosis), and 139 were included in Study 2 (VMA for tobacco and alcohol addiction screening). Patients had to be aged 18 or older, French native speakers, and have sufficient auditory and visual aptitude to interact with the VMA. A more-detailed description of the inclusion and exclusion criteria can be found in ref. ^{6,7}

This project is part of a larger project on virtual reality and clinical phenotyping (PHENOVIRT) that has been approved in compliance with French and European regulations on clinical research by a local ethics committee (Comité pour la Protection des Personnes—Institutional Review Board of University of Bordeaux). All participants gave their written informed consent before entering the study.

VMA description

In study 1, the VMA was developed to conduct a psychiatric interview in order to evaluate MDD according to the Diagnostic and Statistical Manual of Mental Disorder (DSM-5) criteria. In study 2, the agent was developed to screen for current alcohol and tobacco use disorders with an adaptation of the Cigarette Dependence Scale (CDS-5⁴⁹) and the CAGE⁵⁰ questionnaires. The interaction was based on a pre-determined scenario, with several options throughout the case depending on the user's answers but leading to a single end point. The interviews were adapted by sleep specialists and computer scientists to reinforce the credibility and benevolence of the agent, notably by adding small talk and adapting the agent's appearance. The usability of the system and satisfaction with it were considered in the design and pre-tested by the research team.

Both VMAs had a female appearance were displayed on a tablet and talked to the patient with a recorded real voice. The patient could answer the VMA's questions orally thanks to voice recognition. The virtual environment was generated by Unity 3D software (Unity-Technologies, 2014), and gestures were captured by motion capture technology. The software is based on (Fig. 4): (i) a scenario manager, based on decision trees, who coordinates the whole interview and manages the other modules, (ii) a display manager that automatically plays the voice and animations of the virtual human, (iii) an interaction manager, managing speech recognition and the graphical interface to respond to the ECA. A video of our VMA performing an interview can be found in: http://www.sanspy.univ-bordeauxsegalen.fr/Papers/NPJ_Additional_Material.html.

Engagement, acceptance, and trust measures

Patients' perceptions of the VMA were quantified via two questionnaires evaluating trust in the agent and acceptance of the overall system. We used short scales to make it time-efficient and usable in clinical practice, and both questionnaires were answered directly on the tablet after interacting with the VMA.

Acceptance of the E-health system. To measure acceptance of the system, we used the Acceptability E-scale.¹⁵ This self-reported questionnaire assessing acceptance of E-health systems comprises six items (Table 2). The scale comprises a total score and two sub-scores regarding usability (i.e., the perceived ease of using the system) and satisfaction of the device

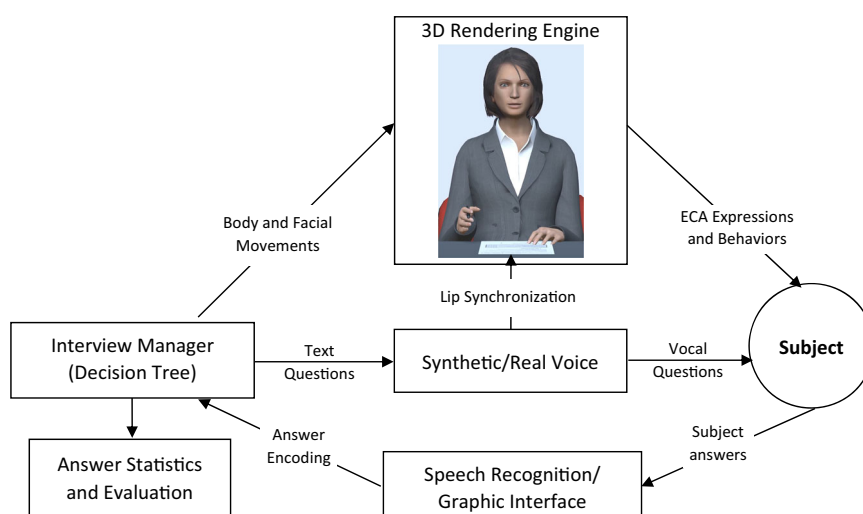


Fig. 4 Architecture of the Embodied Conversational Agent.

Table 2. English and French versions of the ECA Trust Questionnaire (ETQ) and the Acceptability E-scale (AES).

No.	Factor	Item	French version
AES items^{15,19}			
1	Usability	How easy was the computer programme to use?	A quel point avez-vous trouvé ce programme informatique facile d'utilisation?
2	Usability	How understandable were the questions?	A quel point les questions étaient-elles compréhensibles?
3	Satisfaction	How much did you enjoy using this computer programme?	A quel point avez-vous apprécié l'utilisation de ce programme informatique?
4	Satisfaction	How helpful was this computer programme in describing your symptoms and quality of life?	A quel point ce programme informatique vous a-t-il été utile pour décrire vos symptômes et votre qualité de vie?
5	Usability	Was the amount of time it took to complete this computer programme acceptable?	Le temps consacré à répondre à ce programme informatique était-il acceptable?
6	Satisfaction	How would you rate your overall satisfaction with this computer programme?	Comment évaluez-vous votre satisfaction générale de cet outil informatique?
ETQ items			
1	Benevolence	Did you feel that your answers were correctly understood by the virtual agent?	Avez-vous eu l'impression que vos réponses ont bien été comprises par l'agent virtuel?
2	Benevolence	Did you feel that the questions asked by the virtual agent were clear?	Est-ce que les questions posées par l'agent virtuel vous ont paru claires?
3	Benevolence	Did you feel that the interview with the virtual agent was pleasant?	Avez-vous trouvé l'entretien avec l'agent virtuel agréable?
4	Credibility	Would you agree to being cared for by the virtual agent in hospital?	Seriez-vous d'accord que l'agent virtuel participe à votre prise en charge à l'hôpital?
5	Credibility	Would you agree to being cared for by the virtual agent at home?	Seriez-vous d'accord pour que l'agent virtuel participe à votre prise en charge à domicile?
6	Credibility	Did you feel that the virtual agent was competent?	Avez-vous eu l'impression que l'agent virtuel était compétent?

(i.e., the perceived enjoyment of the use and usefulness of the system).^{15,16} Items are based on a five-point Likert scale ranging from 1: "very unsatisfied"; to 5: "very satisfied", where higher scores indicate higher acceptance. This scale was validated in its French version¹⁹ in a previous publication from our team.

Perceived trust in the VMA. Based on the literature,^{20–23} we developed a six-item questionnaire evaluating trust in the VMA (which we named the ECA Trust Questionnaire, Table 2). Items are based on a four-point Likert scale ranging from 0: "Not at all" to 3: "Totally agree", with a total score of 18, higher scores indicating a more favourable attitude toward the agent. The questionnaire is subdivided into two three-item sub-scores regarding perceived credibility (i.e., perception that the agent has the ability and the expertise to conduct a medical intervention) and benevolence of the agent (i.e., perception that the agent is well-intentioned and will accurately take

one's interests into account), each scored out of 9. To validate the psychometric properties of the scale, several statistical analyses were performed. First, to ensure its construct validity, we conducted a principal component analysis. Results validated breaking down the scale into two factors (KMO value: .67), Factor 1 (corresponding to credibility) grouping items 4, 5, and 6; and Factor 2 (corresponding to benevolence) grouping items 1, 2, and 3 (altogether explaining 61.19% of the variance). Varimax-rotated factors loading of the ETQ scale are presented in Supplementary Table 3. To ensure that the ETQ explores factors other than those on the AES scale, we conducted confirmatory factor analyses on the six items of the ETQ scale plus the six items of the AES scale with two latent variables underlying the four sub-scores of ETQ and AES. Root Mean Square Error of Approximation (RMSEA), Standardised Root Mean Square Residual (SRMR) and Comparative Fit Index (CFI), indicated an acceptable model fit (RMSEA = 0.104; SRMR = 0.027; and CFI = 0.988). Second, to ensure the

internal consistency of the ETQ, Cronbach's alpha coefficients were calculated. According to Cronbach's threshold,⁵¹ analyses showed acceptable results (alpha = 0.71, and ranging from 0.66 to 0.71 after removing one item). Finally, floor and ceiling effects measuring the proportions of participants obtaining the lowest and the highest scores for each item were calculated to assess the distribution of responses. Obtained floor effects ranged from 0.6% to 16.0% and ceiling effects ranged from 14.5% to 61.0%. Taken together, these analyses suggest that the scale has satisfactory psychometric properties.

Future engagement with the VMA. To estimate patients' willingness to stay engaged with the VMA during repetitive use, we administered one two-choice question: "Are you willing to engage in a new interaction with the virtual medical agent?" after the interview with the VMA. Patients could answer "Yes" or "No". We refer hereafter to this question as the "Engagement question".

Statistical analyses

Quantitative variables were expressed with means (M), and standard deviations (SD), and qualitative variables were expressed using distributions and percentages. To investigate factors associated with acceptance (usability and satisfaction sub-scores of the AES), trust (credibility and benevolence sub-scores of the ETQ) and engagement (answer to Engagement question), we conducted univariate analyses with Pearson correlation analyses between two continuous variables (age, education, AES sub-scores, ETQ sub-scores), mean comparisons (t test or analysis of variances) for categorical variables (gender, type of sleep disorder, medical domain of the agent) and distribution comparisons (χ^2 tests) when both variables are categorical (when comparison is based on the Engagement question). When associations appeared significant ($p < 0.05$), we conducted multivariate analyses using linear regressions for AES et ETQ sub-scores as dependent variables, and using logistic regressions for Engagement question as the dependent variable.

At last, to identify acceptance and trust thresholds that would induce future patients to engage with the system, we performed ROC analyses using AES and ETQ sub-scores as parameters, and the Engagement question as binary classifier. For these analyses, AUC, sensitivity/specificity and positive/negative predictive reports were presented. A cutoff point was obtained by selecting the point on the ROC curve that maximised both sensitivity and specificity. Analyses were performed using SPSS software (version 18, PASW Statistics), RStudio (version 1.2.1335, RStudio Inc) and MedCalc (version 14.8 for Windows).

Reporting summary

Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 10 July 2019; Accepted: 2 December 2019;

Published online: 07 January 2020

REFERENCES

- Wykes, T. et al. Mental health research priorities for Europe. *Lancet Psychiatry* **2**, 1036–1042 (2015).
- Laranjo, L. et al. Conversational agents in healthcare: a systematic review. *J. Am. Med. Inform. Assoc.* **25**, 1248–1258 (2018).
- Provoost, S., Lau, H. M., Ruwaard, J. & Riper, H. Embodied Conversational Agents in Clinical Psychology: A Scoping Review. *J. Med. Internet Res.* **19**, e151 (2017).
- Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S. & Torous, J. B. Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. *Can. J. Psychiatry* 0706743719828977 (2019). <https://doi.org/10.1177/0706743719828977>.
- Cassell, J. et al. Animated Conversation: Rule-based Generation of Facial Expression, Gesture & Spoken Intonation for Multiple Conversational Agents. in *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques* 413–420 (ACM, 1994). <https://doi.org/10.1145/192161.192272>.
- Philip, P. et al. Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders. *Sci. Rep.* **7**, 42656 (2017).
- Auriacombe, M. et al. Development and validation of a virtual agent to screen tobacco and alcohol use disorders. *Drug Alcohol Depend.* **193**, 1–6 (2018).
- Lucas, G. M. et al. Reporting Mental Health Symptoms: Breaking Down Barriers to Care with Virtual Human Interviewers. *Front. Robot. AI* **4**, 51 (2017).
- Lisetti, C., Amini, R., Yasavur, U. & Riske, N. I Can Help You Change! An Empathic Virtual Agent Delivers Behavior Change Health Interventions. *ACM Trans. Manag. Inf. Syst.* **4**, 1–19 (2013).
- Ren, J., Bickmore, T., Hempstead, M. & Jack, B. Birth Control, Drug Abuse, or Domestic Violence? What Health Risk Topics Are Women Willing to Discuss with a Virtual Agent? in *Intelligent Virtual Agents* (eds. Bickmore, T., Marsella, S. & Sidner, C.) 350–359 (Springer International Publishing, 2014).
- Hopkins, I. M. et al. Avatar Assistant: Improving Social Skills in Students with an ASD Through a Computer-Based Intervention. *J. Autism Dev. Disord.* **41**, 1543–1555 (2011).
- Bickmore, T. W. et al. Managing Chronic Conditions with a Smartphone-based Conversational Virtual Agent. in *Proceedings of the 18th International Conference on Intelligent Virtual Agents - IVA '18* 119–124 (ACM Press, 2018). <https://doi.org/10.1145/3267851.3267908>.
- Henson, P., Wisniewski, H., Hollis, C., Keshavan, M. & Torous, J. Digital mental health apps and the therapeutic alliance: initial review. *BJPsych Open* **5**, e15 (2019).
- Bhugra, D. et al. The WPA- Lancet Psychiatry Commission on the Future of Psychiatry. *Lancet Psychiatry* **4**, 775–818 (2017).
- Tariman, J. D., Berry, D. L., Halpenny, B., Wolpin, S. & Schepp, K. Validation and testing of the Acceptability E-scale for Web-based patient-reported outcomes in cancer care. *Appl. Nurs. Res.* **24**, 53–58 (2011).
- Davis, F. D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Q.* **13**, 319–340 (1989).
- Bevan, N., et al. Standards for Usability, Usability Reports and Usability Measures. in *Human-Computer Interaction. Theory, Design, Development and Practice* (ed. Kurosu, M.) 268–278 (Springer International Publishing, 2016).
- Heerink, M., Kröse, B., Evers, V. & Wielinga, B. Assessing Acceptance of Assistive Social Agent Technology by Older Adults: the Almere Model. *Int. J. Soc. Robot.* **2**, 361–375 (2010).
- Micoulaud-Franchi, J.-A. et al. Validation of the French version of the Acceptability E-scale (AES) for mental E-health systems. *Psychiatry Res.* **237**, 196–200 (2016).
- Cassell, J. & Bickmore, T. External Manifestations of Trustworthiness in the Interface. *Commun. ACM* **43**, 7 (2000).
- Zhang, M., Luo, M., Nie, R. & Zhang, Y. Technical attributes, health attribute, consumer attributes and their roles in adoption intention of healthcare wearable technology. *Int. J. Med. Inf.* **108**, 97–109 (2017).
- Benbasat, I. & Wang, W. Trust In and Adoption of Online Recommendation Agents. *J. Assoc. Inf. Syst.* **6**, 4 (2005).
- Pavlou, P. A. Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with the Technology Acceptance Model. *Int. J. Electron. Commer.* **7**, 101–134 (2003).
- Chen, K. & Chan, A. H. S. Gerontechnology acceptance by elderly Hong Kong Chinese: a senior technology acceptance model (STAM). *Ergonomics* **57**, 635–652 (2014).
- Micoulaud-Franchi, J.-A. et al. Acceptability of Embodied Conversational Agent in a Health Care Context. in *Intelligent Virtual Agents* (eds. Traum, D. et al.) 416–419 (Springer International Publishing, 2016).
- Philip, P., Bioulac, S., Sauteraud, A., Chauton, C. & Olive, J. Could a Virtual Human Be Used to Explore Excessive Daytime Sleepiness in Patients? *Presence Teleoperators Virtual Environ.* **23**, 369–376 (2014).
- Lacour, C. Les diplômés du supérieur en Aquitaine: la région profite de son attractivité (2015).
- Lévy, P. et al. Obstructive sleep apnoea syndrome. *Nat. Rev. Dis. Prim.* **1**, 15015 (2015).
- Durick, J., Robertson, T., Brereton, M., Vetere, F. & Nansen, B. Dispelling Ageing Myths in Technology Design. in *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration* 467–476 (ACM, 2013). <https://doi.org/10.1145/2541016.2541040>.
- Chouvarda, I. G., Goulis, D. G., Lambrinoukaki, I. & Maglaveras, N. Connected health and integrated care: Toward new models for chronic disease management. *Maturitas* **82**, 22–27 (2015).
- World Health Organization. *World Report on Ageing and Health*. (World Health Organization, 2015).
- Suchman, A. L. & Matthews, D. A. What makes the patient-doctor relationship therapeutic? Exploring the connexional dimension of medical care. *Ann. Intern. Med.* **108**, 125–130 (1988).
- Ardito, R. B. & Rabellino, D. Therapeutic Alliance and Outcome of Psychotherapy: Historical Excursus, Measurements, and Prospects for Research. *Front. Psychol.* **2**, 270 (2011).
- Preece, J. et al. *Human-Computer Interaction*. (Addison-Wesley Longman Ltd., 1994).

35. Torous, J. et al. Creating a Digital Health Smartphone App and Digital Phenotyping Platform for Mental Health and Diverse Healthcare Needs: an Interdisciplinary and Collaborative Approach. *J. Technol. Behav. Sci.* **4**, 73–85 (2019).
36. Gracey, L. E. et al. Use of user-centered design to create a smartphone application for patient-reported outcomes in atopic dermatitis. *Npj Digit. Med.* **1**, 33 (2018).
37. Ata, R. et al. Clinical validation of smartphone-based activity tracking in peripheral artery disease patients. *Npj Digit. Med.* **1**, 66 (2018).
38. ISO 9241-210:2019. *Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems.* (2019).
39. National Institute for Health and Care Excellence. NICE digital. NICE <https://www.nice.org.uk/digital>.
40. NHS England. Digital transformation. <https://www.england.nhs.uk/digitaltechnology/>.
41. Medical Research Council. Methods research to support the assessment of diagnostic health technologies. <https://mrc.ukri.org/funding/how-we-fund-research/opportunities/methods-research-to-support-the-assessment-of-diagnostic-health-technologies/> (2018).
42. European Medicines Agency. Big data. <https://www.ema.europa.eu/en/about-us/how-we-work/big-data> (2019).
43. King, M. R. N., Rothberg, S., Dawson, R. & Batmaz, F. Bridging the edtech evidence gap: A realist evaluation framework refined for complex technology initiatives. (2016).
44. Wolverton, C. C. & Lanier, P. A. Utilizing the Technology-Organization-Environment Framework to Examine the Adoption Decision in a Healthcare Context. *Handb. Res. Evol. IT Rise E-Soc.* 401–423 (2019). <https://doi.org/10.4018/978-1-5225-7214-5.ch018>.
45. Sittig, D. F. & Singh, H. A New Socio-technical Model for Studying Health Information Technology in Complex Adaptive Healthcare Systems. *Qual. Saf. Health Care* **19**, i68–i74 (2010).
46. Torous, J. B. et al. A Hierarchical Framework for Evaluation and Informed Decision Making Regarding Smartphone Apps for Clinical Care. *Psychiatr. Serv.* **69**, 498–500 (2018).
47. Torous, J. et al. The Emerging Imperative for a Consensus Approach Toward the Rating and Clinical Recommendation of Mental Health Apps. *J. Nerv. Ment. Dis.* **206**, 662–666 (2018).
48. Venkatesh, V. & Davis, F. D. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal. *Field Stud. Manag. Sci.* **46**, 186–204 (2000).
49. Etter, J.-F., Houezec, J. L. & Perneger, T. V. A Self-Administered Questionnaire to Measure Dependence on Cigarettes: The Cigarette Dependence Scale. *Neuropsychopharmacology* **28**, 359 (2003).
50. Bush, B., Shaw, S., Cleary, P., Delbanco, T. L. & Aronson, M. D. Screening for alcohol abuse using the cage questionnaire. *Am. J. Med.* **82**, 231–235 (1987).
51. Cronbach, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika* **16**, 297–334 (1951).

ACKNOWLEDGEMENTS

This project was supported by the national grants LABEX BRAIN (ANR-10-LABX-43), EQUIPEX PHENOVIRT (ANR-10-EQPX-01), and IDEX (ANR-10-IDEX-03-02). We thank the

outpatients at the Sleep Clinic who took time to test our virtual medical agent and gave us precious feedback.

AUTHOR CONTRIBUTIONS

L.D., J.A.M., and P.P., wrote the manuscript. L.D. performed the statistical analyses on the collected data. E.D.S. developed and tested the virtual medical agent. J.A.M., A.S., M.A., and P.P., conceived and designed the study. F.S. and A.S., ran the protocol and acquired all the data for the study. P.P. is the principal investigator in charge of the study.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at <https://doi.org/10.1038/s41746-019-0213-y>.

Correspondence and requests for materials should be addressed to P.P.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020