

Case study in choosing a graphical character to support reminiscence therapy for those living with dementia

Matthew P. Aylett
Heriot-Watt University, CereProc Ltd.
Edinburgh, UK
matthewa@cereproc.com

Katerina Pappa
University of Strathclyde
Glasgow, UK
katerina.pappa@strath.ac.uk

Mei Yii Lim
Heriot-Watt University
Edinburgh, UK
M.Lim@hw.ac.uk

Bruce W. Wilson
Heriot-Watt University
Edinburgh, UK
bww1@hw.ac.uk

Mario Parra
University of Strathclyde
Glasgow, UK
mario.parra-rodriguez@strath.ac.uk

Ruth S. Aylett
Heriot Watt University
Edinburgh, UK
R.S.Aylett@hw.ac.uk

ABSTRACT

The present study describes the process employed to identify a suitable intelligent virtual agent (IVA) for the AMPER App supporting reminiscence therapy for those living with dementia through the use of a facilitating agent. This included three distinct phases: 1) co-creation with project stakeholders and identification of a set of desirable IVA traits; 2) a blind internal team rating process to select a subset from available IVAs; 3) a character survey with healthy older adults to select a final 4 IVAs (2 male, 2 female). We analyse the results, assess inter-subject agreement, and suggest guidelines for IVA selection.

CCS CONCEPTS

• **Human-centered computing** → **Interactive systems and tools**; • **Applied computing** → **Health informatics**.

KEYWORDS

social agents, dementia

ACM Reference Format:

Matthew P. Aylett, Katerina Pappa, Mei Yii Lim, Bruce W. Wilson, Mario Parra, and Ruth S. Aylett. 2024. Case study in choosing a graphical character to support reminiscence therapy for those living with dementia. In *ACM International Conference on Intelligent Virtual Agents (IVA '24)*, September 16–19, 2024, GLASGOW, United Kingdom. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3652988.3673964>

1 INTRODUCTION

Intelligent virtual agents (IVAs), applied in a variety of fields, e.g.: entertainment [25], education [11], museums [17], coaching [33] and healthcare [3, 31] are usually designed to interact with humans. These are social environments, so that such IVAs require a natural interaction interface to maintain long-term engagement with interaction partners. Including IVAs can increase user satisfaction [33], make a technology more accessible [3], and offer increased functionality [12]. However, as Aylett et al. point out [2], "When we

build systems that harness the power of human conversation we are dealing with something that is central to a human sense of self". Thus, in designing an IVA, we enter the social world and issues such as gender, ethnicity and diversity become relevant [7].

The AMPER (Agent-based Memory Prosthesis to Encourage Reminiscing) project aims to develop an app with a story facilitator agent, based on autobiographical memory [28], that supports reminiscing between a user living with dementia and their carer. Autobiographical life stories are important not only for human social interaction but for engagement with artificial agents [5, 10]. Such agents require *human-like* memory processes to meet user expectations of life-likeness, intelligence and responsiveness. [8, 19, 21, 32].

The use of an embodied agent with this vulnerable user group aims to be supportive and to increase engagement (a significant challenge in apps designed to support people living with dementia [1, 23]). The choice of agent appearance and voice is thus critical.

In this paper we first share the process we used to select the IVA for AMPER adopting the co-creation design approach. Project stakeholders and healthy older adults were actively involved early in this selection process consistent with principles of user-centred design and co-creation [20, 24]. We assess its strengths and weaknesses and what we learned from applying it. Secondly, we briefly discuss how our results relate to the issue of avoiding ageist stereotypes.

2 PREVIOUS WORK

IVAs are well-suited to healthcare domains. They are not intended to replace human healthcare providers but to enable them to focus on exceptional cases by automating routine parts of their work. Such characters may be accessed at any time of day allowing patients to interact at their own convenience and pace [3]. Thus, in a study of an IVA-based discharge nurse, patients preferred receiving their discharge instructions from the IVA than their doctor or nurses because they did not feel rushed [37]. In a study of patients with chronic conditions, those interacting with a character on a smartphone reported a significantly higher quality of life at the end of the intervention period compared to those undergoing standard care [6]. IVAs can not only deliver practical medical information to patients but also social, emotional and relational messages. They can establish trusting relationships with users to promote behaviour change interventions [4].

One challenge in IVA selection is that engineers may feel they can make choices without input from potential users or stakeholders. However, unconscious bias is a powerful driving force in forming

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

IVA '24, September 16–19, 2024, GLASGOW, United Kingdom

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0625-7/24/09

<https://doi.org/10.1145/3652988.3673964>

the criteria for what is *appropriate*. For example in Richards et al.'s [31] study involving co-design of an IVA for stroke victims, the IVA gender was chosen up front because "*females are associated with empathetic caring dialogue*". However, the hairstyle, face shape and whether the IVA had glasses were then chosen based on team input and that of stroke survivors. This highlights the difficulty in IVA creation: some assumptions must be made, but which should be part of the design process, and which should not?

In the effect known as similarity-attraction, some studies [27, 34, 36] show humans lean towards characters that have similar traits, gender, and age to themselves, finding them not only more attractive, but also more reliable, trustworthy, and knowledgeable. However, other studies contradict this, showing more of a complementary-attraction effect: a preference for characters that complement their personality [18]. Other work [14, 22] found that whether humans trust and respond favourably to robots whose personalities are similar or complementary to their own depends on the interaction context and the robot role. Therefore, it is important to take context into consideration - the intended role of the character and the target audience - when designing an appropriate appearance for a graphical character [7]. Moreover, the most usable interfaces are those that can be modified by the user, with user satisfaction increasing when the interface is personalised [35]. This study describes the co-creation process employed for the purposes of identifying a suitable IVA specific to the AMPER project.

3 A CO-DESIGN SELECTION PROCESS

3.1 Stakeholder workshop

Our first step was an informal workshop with the AMPER project stakeholder group consisting of N=9 individuals including ReD-Lat (Multi-Partner Consortium to Expand Dementia Research in Latin America) members, Neuroprogressive and Dementia Research Network (NDRN) members, Alzheimer's Scotland National Dementia Carers Action Network (NDCAN) members and industry experts in memory/memorabilia preservation such as After Cloud.

Stakeholders were asked to provide feedback on a small set of character exemplars. These covered a variety of potential character designs: Cartoon/Naturalistic, 2D/3D, Younger/Older, Human/Non-human. Each character was assessed by N=3-4 stakeholders based on the following questions:

- (1) Give up to 5 descriptive words about this character.
- (2) What is good about it?
- (3) What is bad about it?
- (4) Do you think this is a good choice as a storyteller for our specific target group WHY? Or WHY NOT?
- (5) What would you keep?
- (6) What would you change?

A moderated discussion then took place on types of possible characters and what traits might be important for the AMPER storyteller character. Outcomes were: 1) the stakeholders preferred a cartoon 3D, Older, Human-like, character; 2) the stakeholders identified the traits of friendliness, kindness, knowledgeability, amicability, trustworthiness, and non-stereotypical appearance as important for the digital character.

3.2 Internal character assessment

The main aim of internal character assessment was to reduce the number of characters for final external ratings. First, an AMPER

team member identified seventeen potential commercially-available characters (N = 7 females) based on price, format and appearance criteria gathered from the stakeholder workshop. All team members (n=5) then individually rated each against the traits of friendliness, kindness, knowledge, amicability, trustworthiness, and non-stereotypical appearance established by the stakeholders. A character scored 1 if thought to exhibit a specific trait and 0 if thought to lack it. A total and an average trait score were calculated for each character. Raters were blind to each other's scores. Six characters were immediately rejected due to receiving the lowest rating. The remaining eleven were further assessed with another three excluded due to high price or animation complexity. Four female and four male characters were deemed suitable for the next stage.

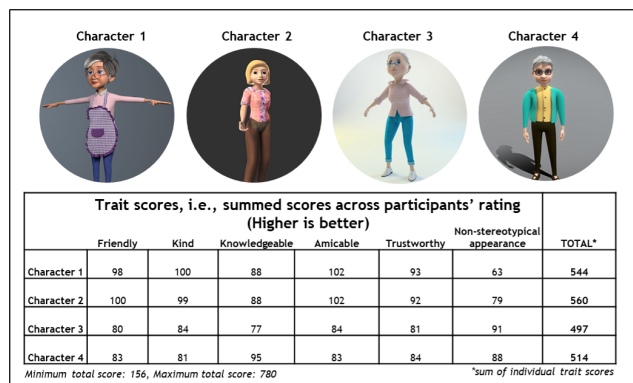


Figure 1: Trait scores for the 4 short listed female IVAs.

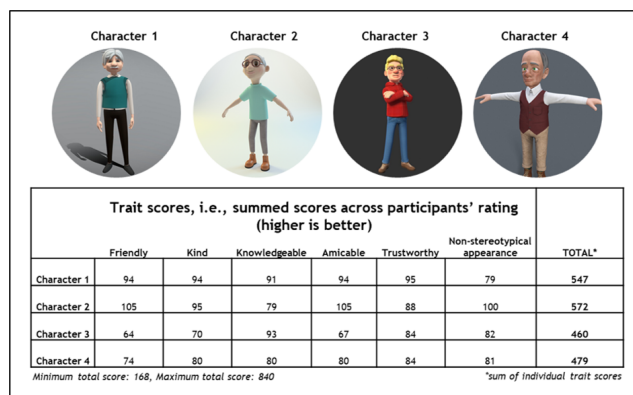


Figure 2: Trait scores for the 4 short listed male IVAs.

3.3 External Character Selection

These eight characters were rated by healthy older adults. 26 participants (age range 59-87) completed the female survey and 28 participants (age range 59-87) completed the male survey, answering eight questions. Most participants completed both surveys: the order was counterbalanced. Question 1 asked participants to write the first three things that came to mind when viewing the character. Questions 2-7 asked participants to rate characters against the traits identified by the stakeholders in the first step, i.e., friendliness, kindness, knowledge, amicability, trustworthiness, and non-stereotypical appearance using a 5-point Likert agreement scale (from 1-Strongly disagree to 5-Strongly agree). Question 8 asked participants to freely write any other comments and/or feedback they had. For each character, the total score per trait was calculated

by summing up the individual score across participants and the grand total score was calculated by summing up the total score across traits across participants. This gave a final selection of two male and two female characters: see Fig. 1 and Fig. 2 for scores.

4 RESULTS

Statistical analysis was performed only for the external character selection.

4.1 Correlation between Features

Statistical analyses as well as descriptive measures for the characters were produced using R statistical software (v4.2.3, R Core Team 2023). Spearman Rank correlation analyses were performed to examine associations between the different character traits. False Discovery Rates (FDR) were applied to correct for multiple comparisons [29]. Analysis shows a strong correlation between *Friendly/Kind/Amicable* (Spearman’s rho 0.6 – 0.79). These traits have weak correlation with *Knowledgeable* (Spearman’s rho < 0.25), while *Trustworthy* shows a moderate correlation (Spearman’s rho 0.3 – 0.4) with other traits. These are all positive. The exception is *Non-stereotypical* which does not correlate with any other traits. We can regard the traits as forming three groups: *Friendly/Amicable/Kind*; *Knowledgeable/Trustworthy*; and *Non-stereotypical*.

Table 1: Correlation between traits: Spearman rho (2 dp) (p-value). frd - Friendly, knd - Kind, knw - Knowledgeable, amc - Amicable, trs - Trustworthy, nst - Non-stereotypical.

	frd	knd	knw	amc	trs
frd	-	-	-	-	-
knd	0.66 (0.00)	-	-	-	-
knw	0.17 (0.01)	0.20 (0.00)	-	-	-
ami	0.70 (0.00)	0.67 (0.00)	0.16 (0.02)	-	-
trs	0.37 (0.00)	0.47 (0.00)	0.37 (0.00)	0.44 (0.00)	-
nst	0.04 (ns)	-0.04 (ns)	0.01 (ns)	-0.00 (ns)	0.06 (ns)

4.2 Agreement

Agreement was calculated for ratings given in the external selection. Intraclass Correlation Coefficient (ICC) estimates and their 95% confidence intervals were calculated using R statistical software (v4.2.3, [30]) based on a mean-rating ($k = 27$), absolute-agreement, one-way random-effects model.

A one-way random-effects model was chosen since not all participants completed both surveys and there were cases of missing data. The analysis was computed using the function `iccNA` from package `irrNA` [9] which works on (unbalanced) incomplete datasets without any imputation of missing values. An ICC analysis was computed on the summed score across the six character traits for $N=8$ characters ($N=4$ females) for each rater. The estimated ICC = 0.88 ($p < .001$) with 95% confidence interval = 0.73-0.97, suggesting good agreement [16] between raters.

5 DISCUSSION

Trait correlation findings suggest that the character traits are not orthogonal. Standard trait markup systems (such as Big 5 [26]) apply complex processes to discover underlying orthogonal factors, directly relate these to trait names and finally validate a questionnaire. Furthermore, the stakeholders identified three traits which heavily correlated with each other, i.e., *Friendly*, *Kind* and *Amicable*, which could be interpreted as an informal weighting. In effect, appearing

Friendly proved three times more important than being perceived as knowledgeable, trustworthy or non-stereotypical. Stakeholders may expect such features in agents/characters that will interact with patients with dementia [13]) but this needs further research.

The *Trustworthy* and *Knowledgeable* traits show a relationship with the friendly traits, with some divergence. This reflects previous findings in the literature revealing a tension between being authoritative (also contributing to a sense of trust and competence) and being friendly [15, 34]. In contrast, the *Non-stereotypical* trait showed no relationship with any of the other traits and was identified as an outlier compared to the others.

Furthermore, rating agreement was high amongst healthy older adults, suggesting similarities in the way raters perceived the different character traits and IVAs. This could also be influenced by the fact that three of the five traits heavily correlated, i.e., *Friendly*, *Kind*, *Amicable*.

5.1 IVA Selection Guidelines

Given the results of our process, we can suggest a set of guidelines for future evaluation of graphical characters in a co-design context:

- An informal set of traits collected from stakeholders plus a two stage process of selecting characters (rating within the team; rating with healthy control users of similar age, background to the target user) is an appropriate basis for selection and leads to good agreement among raters.
- Rather than use a trait for how stereotypical a character may look, we advise indirect non-bias questions to gather their subjective opinions such as “*What do you like/dislike about this character, What would you retain/eliminate from the character? And why?*”
- If pose and dress cannot be controlled, consider rating the character on head appearance only or explicitly ask for feedback on pose and dress.
- Allow users to choose more than one character. In this work, for both male and female, one selected character is more traditional and the other less so. This supports both similarity-attraction and the complementary-attraction effect offering users not just a choice of gender but also some variation in character style.

5.2 Limitations

The present study has certain limitations. First, the characters included in the external character selection were commercially available and identified by the AMPER team rather than generated from scratch using automated tools. Second, the fact that some participants completed two surveys while others only completed one survey, might have introduced potential bias to the data. Finally, the present study is only a small scale evaluation and therefore, the findings should be interpreted with caution as they are relevant only for the purposes of the AMPER study.

5.3 Conclusions

To conclude, our co-creation can be used as a guide for IVA selection through involving stakeholders early in the process rather than researchers simply selecting the agent themselves. Nevertheless, a co-design process is resource-limited, both by available research effort, and also by how much time stakeholders can commit. Therefore, a co-creation process should be flexible and adaptable.

ACKNOWLEDGMENTS

Support for this research was provided by EPSRC (AMPER: award number EP/V056131/1, EP/V05564X/1 and Designing Conversational Assistants to Reduce Gender Bias: award number EP/T023767/1).

REFERENCES

- [1] Astell AJ, Bouranis N, Hoey J, Lindauer A, Mihailidis A, Nugent C, and Robillard JM. 2019. Technology and Dementia: The Future is Now. *Dement Geriatr Cogn Disord*. 47, 3 (2019), 131–139. <https://doi.org/10.1159/000497800>
- [2] Matthew Peter Aylett and Marta Romeo. 2023. You Don't Need to Speak, You Need to Listen: Robot Interaction and Human-Like Turn-Taking. In *Proceedings of the 5th International Conference on Conversational User Interfaces*. 1–5.
- [3] Timothy Bickmore. 2022. *Health-Related Applications of Socially Interactive Agents* (1 ed.). Association for Computing Machinery, New York, NY, USA, 403–436. <https://doi.org/10.1145/3563659.3563672>
- [4] Timothy Bickmore and Rosalind Picard. 2004. Towards caring machines. 1489–1492. <https://doi.org/10.1145/985921.986097>
- [5] Timothy Bickmore, Daniel Schulman, and Langxuan Yin. 2009. Engagement vs. deceit: Virtual humans with human autobiographies. In *Intelligent Virtual Agents: 9th International Conference, IVA 2009 Amsterdam, The Netherlands, September 14–16, 2009 Proceedings 9*. Springer, 6–19.
- [6] Timothy W. Bickmore, Everlyne Kimani, Ha Trinh, Alexandra Pusateri, Michael K. Paasche-Orlow, and Jared W. Magnani. 2018. Managing Chronic Conditions with a Smartphone-based Conversational Virtual Agent. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents* (Sydney, NSW, Australia) (IVA '18). Association for Computing Machinery, New York, NY, USA, 119–124. <https://doi.org/10.1145/3267851.3267908>
- [7] Marissa Bond, David M. W. Powers, and Parimala Raghavendra. 2021. *Creating Engaging Embodied Conversational Agents*. Springer Singapore, Singapore, 45–60. https://doi.org/10.1007/978-981-16-2312-7_6
- [8] Cyril Brom, Jiří Lukavský, and Rudolf Kadlec. 2010. Episodic memory for human-like agents and human-like agents for episodic memory. *International Journal of Machine Consciousness* 2, 02 (2010), 227–244.
- [9] M Brueckl and F Heuer. 2022. irrNA: Coefficients of Interrater Reliability - Generalized for Randomly Incomplete Datasets. (2022). <https://CRAN.R-project.org/package=irrNA>
- [10] J Campos. 2010. *MAY: my Memories Are Yours. An interactive companion that saves the users memories*. Ph. D. Dissertation. Master thesis, Instituto Superior Técnico.
- [11] M. Core, D. Traum, H.C. Lane, W. Swartout, J. Gratch, M. van Lent, and S. Marsella. 2006. Teaching Negotiation Skills through Practice and Reflection with Virtual Humans. *Simulation* 82, 11 (2006), 685–701.
- [12] Kerstin Dautenhahn. 1998. Story-Telling in Virtual Environments. <https://api.semanticscholar.org/CorpusID:2869853>
- [13] Anna Hunt, Anthony Martyr, Laura D Gamble, Robin G Morris, Jeanette M Thom, Claire Pentecost, and Linda Clare. 2023. The associations between personality traits and quality of life, satisfaction with life, and well-being over time in people with dementia and their caregivers: findings from the IDEAL programme. *BMC geriatrics* 23, 1 (2023), 1–12.
- [14] M. Joosse, M. Lohse, J. G. Pérez, and V. Evers. 2013. What you do is who you are: The role of task context in perceived social robot personality. In *Proc. of Robotics and automation 2013 IEEE international conference*. IEEE, Karlsruhe, Germany, 2134–2139. <https://doi.org/10.1109/ICRA.2013.6630863>
- [15] Reshmashree B Kantharaju, Dominic De Franco, Alison Pease, and Catherine Pelachaud. 2018. Is two better than one? Effects of multiple agents on user persuasion. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. 255–262.
- [16] Terry K Koo and Mae Y Li. 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine* 15, 2 (2016), 155–163.
- [17] S. Kopp, L. Gesellensetter, N. C. Krämer, and I. Wachsmuth. 2005. A Conversational Agent as Museum Guide - Design and Evaluation of a Real-World Application. In *Intelligent Virtual Agents Conference*. Springer-Verlag, Greece, 329–343.
- [18] Kwan Min Lee, Wei Peng, Seung-A Jin, and Chang Yan. 2006. Can Robots Manifest Personality?: An Empirical Test of Personality Recognition, Social Responses, and Social Presence in Human-Robot Interaction. *J. of Commun.* 56 (2006), 754–772. Issue 4. <https://doi.org/10.1111/j.1460-2466.2006.00318.x>
- [19] Iolanda Leite, Carlos Martinho, and Ana Paiva. 2013. Social robots for long-term interaction: a survey. *International Journal of Social Robotics* 5 (2013), 291–308.
- [20] Veiko Lember, Taco Brandsen, and Piret Tõnurist. 2019. The potential impacts of digital technologies on co-production and co-creation. *Public Management Review* 21, 11 (2019), 1665–1686.
- [21] Mei Yui Lim. 2012. Memory models for intelligent social companions. *Human-Computer Interaction: The Agency Perspective* (2012), 241–262.
- [22] Mei Yui Lim, José David Aguas Lopes, David A. Robb, Bruce W. Wilson, Meriam Moujahid, Emanuele De Pellegrin, and Helen Hastie. 2022. We are all Individuals: The Role of Robot Personality and Human Traits in Trustworthy Interaction. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. 538–545. <https://doi.org/10.1109/RO-MAN53752.2022.9900772>
- [23] Petra Maresova, Signe Tomsone, Petre Lameski, Joana Madureira, Ana Mendes, Eftim Zdravevski, Ivan Chorbev, Vladimir Trajkovik, Moriah Ellen, and Kasper Rodile. 2018. Technological solutions for older people with Alzheimer's disease. *Current Alzheimer Research* 15, 10 (2018), 975–983.
- [24] Patrizia Marti and Liam J Bannon. 2009. Exploring user-centred design in practice: Some caveats. *Knowledge, technology & policy* 22 (2009), 7–15.
- [25] M. Mateas and A. Stern. 2003. Facade: An Experiment in Building a Fully-Realized Interactive Drama. In *Game Developers Conference: Game Design Track*. San Jose, California.
- [26] Robert R McCrae and Paul T Costa. 1987. Validation of the five-factor model of personality across instruments and observers. *Journal of personality and social psychology* 52, 1 (1987), 81.
- [27] Clifford Nass, Katherine . Isbister, and Eun-Ju Lee. 2001. *Truth is beauty: researching embodied conversational agents*. MIT Press, Cambridge, MA, USA, 374–402.
- [28] Katherine Nelson. 1993. The psychological and social origins of autobiographical memory. *Psychological science* 4, 1 (1993), 7–14.
- [29] Nathan Pike. 2011. Using false discovery rates for multiple comparisons in ecology and evolution. *Methods in ecology and Evolution* 2, 3 (2011), 278–282.
- [30] R R Core Team et al. 2013. R: A language and environment for statistical computing. (2013).
- [31] Deborah Richards, Paulo Sergio Miranda Maciel, and Heidi Janssen. 2023. The Co-Design of an Embodied Conversational Agent to Help Stroke Survivors Manage Their Recovery. *Robotics* 12, 5 (2023), 120.
- [32] Caroline Rizzi, Colin G Johnson, Fabio Fabris, and Patricia A Vargas. 2017. A situation-aware fear learning (safel) model for robots. *Neurocomputing* 221 (2017), 32–47.
- [33] Dylan GM Schouten, Agnes A Deneka, Mariët Theune, Mark A Neerinx, and Anita HM Cremers. 2023. An embodied conversational agent coach to support societal participation learning by low-literate users. *Universal access in the information society* 22, 4 (2023), 1215–1241.
- [34] Silke ter Stal, Monique Tabak, Harm op den Akker, Tessa Beinema, and Hermie Hermens. 2020. Who do you prefer? The effect of age, gender and role on users' first impressions of embodied conversational agents in eHealth. *International Journal of Human-Computer Interaction* 36, 9 (2020), 881–892.
- [35] J. Ventura, M. Ventura, and J.C. Olabe. 2005. Embodied conversational agents: developing usable agents [education instruction applications]. In *Proceedings. IEEE SoutheastCon, 2005*. 663–669. <https://doi.org/10.1109/SECON.2005.1423322>
- [36] Steve Whittaker, Yvonne Rogers, Elena Petrovskaya, and Hongbin Zhuang. 2021. Designing Personas for Expressive Robots: Personality in the New Breed of Moving, Speaking, and Colorful Social Home Robots. *ACM Trans. Human-Robot Interact.* 10, 8 (February 2021), 25 pages. Issue 1. <https://doi.org/10.1145/3424153>
- [37] Shuo Zhou, Timothy Bickmore, Michael Paasche-Orlow, and Brian Jack. 2014. Agent-User Concordance and Satisfaction with a Virtual Hospital Discharge Nurse. In *Intelligent Virtual Agents*, Timothy Bickmore, Stacy Marsella, and Candace Sidner (Eds.). Springer International Publishing, Cham, 528–541.