Virtual Assistants are Unlikely to Reduce Patient Non-Disclosure

Corinne Jorgenson[1], Ali I. Ozkes[2], Jurgen Willems[3], Dieter Vanderelst[1]

[1] University of Cincinnati
[2] SKEMA Business School, GREDEG, Université Côte d'Azur
[3] Vienna University of Economics and Business
{jorgencj,vanderdt}@ucmail.uc.edu, jurgen.willems@wu.ac.at, ali.ozkes@skema.edu

Abstract

The ethical use of AI typically involves setting boundaries on its deployment. Ethical guidelines advise against practices that involve deception, privacy infringement, or discriminatory actions. However, ethical considerations can also identify areas where using AI is desirable and morally necessary. One area where ethical considerations can make AI deployment imperative is healthcare. For example, patients often withhold pertinent details from healthcare providers due to fear of judgment. However, utilizing virtual assistants to gather patients' health histories could be a potential solution. Ethical imperatives support using such technology if patients are more inclined to disclose information to an AI system. This article presents findings from several survey studies investigating whether virtual assistants can reduce non-disclosure behaviors. Unfortunately, the evidence suggests that virtual assistants are unlikely to minimize non-disclosure. Therefore, the potential benefits of virtual assistants due to reduced non-disclosure are unlikely to outweigh their ethical risks.

Introduction

The use of virtual assistants (also called conversational agents or virtual humans) is expanding rapidly (Del Valle, Llorca Albareda, and Rueda 2024), including in healthcare applications (Laranjo et al. 2018). Milne-Ives et al. (2020) reviewed existing virtual assistant healthcare care applications. They found that virtual assistants are used in various areas, including clinical decision- or triage support, screening or diagnosis, and education. One motivation for integrating virtual assistants into healthcare is to mitigate existing gaps and meet unaddressed requirements in healthcare services (Luxton 2020). Additional advantages include improving access to care or improving its quality.

The development of virtual agents for healthcare also raises growing concerns about ethical issues (See Luxton 2020; Del Valle, Llorca Albareda, and Rueda 2024; Luxton 2014, for overviews). For example, Luxton (2020) identified four potential issues with the use of virtual assistants in healthcare: bias, unequal access, risk of harm, and privacy concerns (See also Luxton and Watson 2023). Emerging design guidelines might mitigate this (Piñeiro Martín et al. 2022; Luxton 2014).

Ethics are often considered to constrain and limit emerging technologies by banning specific applications (See Abrams 2024, for examples). However, as Coeckelbergh (2020) argues, ethics can also be positive and supportive, encouraging technology development guided by what we consider important as a society. The ethical risks incurred by virtual assistants have received substantial attention. In contrast, their potential ethical benefits have been less explored. Nevertheless, the benefits of virtual assistants in healthcare could make their use imperative and ethically desirable. For example, Luxton (2020) mentions that virtual assistants might be perceived as free from personal bias. Therefore, users might feel less anxious when discussing sensitive issues such as risk behaviors with a conversational agent. This could reduce patient non-disclosure by avoiding potential judgment, a common reason people withhold information from their healthcare providers (Levy et al. 2018). At the same time, virtual assistants could increase the sense of anonymity while maintaining a sense of connection with patients, which could also reduce psychological barriers (Lucas et al. 2014).

In this paper, we present several survey studies to understand whether the use of virtual assistants could increase people's comfort with disclosing health information and reduce non-disclosure. In the discussion, we compare our data and methodology with previous work.

Experiment 1

Experiment 1 is conceived as a manipulation check. This simple online experiment assessed whether participants would respond discriminatorily when rating their comfort and truthfulness in providing a doctor with information on different aspects of their health history. In our studies, participants are not actually in the described situation, nor do we ask them to remember a specific past event. Therefore, it is important to assess whether presenting respondents with a hypothetical scenario makes them perceive that providing some information might be more sensitive or stigmatizing.

Methods Experiment 1

In Experiment 1, we presented participants with a scenario asking them to imagine that they were visiting a new doctor for the first time. We told them that this doctor would be

their primary care provider, addressing their general physical and mental health concerns. The doctor's office must collect their health history to establish their healthcare record as a new patient. The categories of information that the doctor's office must collect include individual medical history, surgical history, family medical history, social history, allergies, and medications they are taking. We asked participants to rate their comfort in providing information on these six categories using a slider. The prompt was the following.

> Imagine that you are visiting a new doctor for the first time. You expect that this doctor will be your primary care provider, addressing your general physical and mental health concerns as they arise.
>
> During your first visit, your doctor's office will need to collect your health history as part of establishing your healthcare record as a new patient. The categories of information your doctor's office needs to collect include your individual medical history, past surgical history, family medical history, social history, allergies, and medications you are taking or may have recently stopped taking.
>
> Using the sliders below, please indicate your general comfort level providing each category of information.

The following texts were provided as descriptions of the six categories of health history.

1. Medical history (e.g., activity level, chronic conditions like diabetes, etc.)
2. Surgical history (e.g., tonsillectomy)
3. Family medical history (e.g., family member with high blood pressure or mental illness)
4. Social history (e.g., sexual behavior, recreational drug use, risk behaviors)
5. Allergies and insensitivities (e.g., nut allergy, gluten intolerance)
6. Medications you are taking or have recently stopped taking (e.g., over-the-counter pain relief, prescription antibiotics)

After providing their comfort ratings, we showed a new page of the survey on which the participants were asked to rate how likely they were to provide complete and truthful information to the doctor when taking the health history for each of the six categories of health information. In particular, the prompt used was the following.

> On a scale from 'not at all likely' to 'completely likely,' how likely are you to provide truthful and complete information to your doctor when they take your health history?

As part of the instructions, we told participants that they would be asked a question about their preferred healthcare provider. Participants were instructed to select 'Nurse' when asked the question. Therefore, this served as an attention question, testing whether the participants had read the instructions. At the end of the survey, we invited participants to complete several demographic questions.

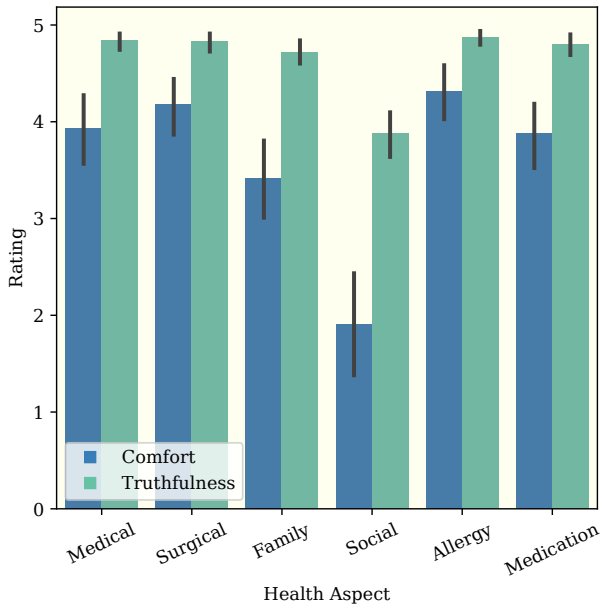

Figure 1: Results of Experiment 1. The graph shows the average comfort in providing information about each of the six aspects of health history. The graph also shows the intention to respond truthfully to questions regarding each category.

We recruited undergraduate students at the University of Cincinnati who participated for course credit. The experiments reported in this paper were approved by the University of Cincinnati IRB (2022-0957).

Results Experiment 1

In total, 181 people completed the survey. Of these, we retained 113 respondents who correctly completed the attention question. The demographics of the participants were as follows. Age: 18 (25%), 19 (53%), 20 (18%), 21 (3%), 22 (1%) Ethnicity: Asian (6%), Black or African American (8%), Other (10%), White (76%) Gender: Female (80%), Male (18%), Non-binary (3%).

Figure 1 shows the results for Experiment 1. The graph shows the respondents indicated they were generally comfortable providing health history information to the doctor's office. The *Social* category was associated with the lowest level of comfort. The responses for the likelihood of responding truthfully to a doctor's questions mirrored this pattern of results. The *Social* category was associated with the lowest reported likelihood to answer completely and truthfully.

Table 1 shows the results of running a mixed linear model with the participants as a random variable fitting the reported comfort level as a function of the six categories of health history. Including participants as random effects allows the model to account for this within-subject correla-

	Coef.	Std. Err.	z	P> |z|
Intercept	1.912	0.197	9.686	0.000
T.1	2.018	0.211	9.570	0.000
T.2	2.265	0.211	10.745	0.000
T.3	1.504	0.211	7.135	0.000
T.5	2.407	0.211	11.417	0.000
T.6	1.965	0.211	9.318	0.000
Group Var	1.889	0.213		

Table 1: Experiment 1, Comfort rating

	Coef.	Std. Err.	z	P> |z|
Intercept	3.876	0.066	58.610	0.000
T.1	0.965	0.078	12.313	0.000
T.2	0.956	0.078	12.200	0.000
T.3	0.841	0.078	10.732	0.000
T.5	1.000	0.078	12.765	0.000
T.6	0.929	0.078	11.861	0.000
Group Var	0.147	0.051		

Table 2: Experiment 1, Truth level

tion or variability. We used the *Social* category as the reference in a *Treatment* contrast. This analysis revealed that participants reported being significantly more comfortable discussing each category compared to the *Social* category ($p < 0.01$).

Table 2 reports the results of a mixed linear model, fitting the self-reported level of truthfulness using the different levels of health history data as a predictor. Again, we used the *Social* category as a reference level. We found that participants indicated they would be significantly less truthful in this category than in others.

Discussion Experiment 1

Experiment 1 confirmed that the health categories used in our survey elicited different comfort levels when revealing information to a healthcare professional. In particular, we found that the *Social* category resulted in the lowest comfort levels and self-reported intention of disclosing complete and truthful information. Therefore, the results of Experiment 1 can be considered as a successful manipulation check. Despite us asking participants about a hypothetical scenario, their responses indicate that they perceived some information as potentially more sensitive or stigmatizing (i.e., the *Social* aspects of their health history).

Experiment 2

Methods Experiment 2

The methodology for Experiment 2 was identical to that of Experiment 1, except that we asked the participants different questions. As in Experiment 1, we asked participants to rate how comfortable they would be disclosing six different kinds of health history information to a doctor. After providing their comfort ratings, we asked the participants whether, if given a choice, they would rather have a doctor or virtual assistant collect each category of health history. Hence, Experiment 2 is the first of our studies to assess whether virtual assistants could be ethically imperative because they reduce non-disclosure of sensitive or stigmatizing information by assessing participants' preferences for a doctor or a virtual assistant. We recruited undergraduate students at the University of Cincinnati who participated for course credit.

Results Experiment 2

We recruited 132 participants who completed the survey. Of these, 77 correctly answered the attention question. We analyzed the data of these participants. Their demographics were as follows. Age: 18 (18%), 19 (42%), 20 (28%), 21 (4%), 22 (6%), 23 (1%). Ethnicity: Asian (10%), Black or African American (5%), Other (6%), White (78%). Gender: Female (70%), Male (27%), Non-binary (3%).

Figure 2a presents the comfort levels reported for each category of health history information. As in Experiment 1, we found that the comfort levels differed between categories, with the *Social* category eliciting the lowest reported comfort levels. Figure 2b shows whether the participants preferred that their information be collected by a doctor (negative values) or a virtual assistant (positive values).

The comfort levels followed the same pattern as in Experiment 1 (Figure 2a). Participants were less comfortable providing information in the category *Social*. Overall, participants preferred that their health history be taken by a doctor (negative end of the scale, Figure 2b). However, their preference was least outspoken for the *Social* category (results of linear model not shown due to space limitations). Indeed, a post hoc contrast test indicated that for this category, the participants' reported preference did not significantly differ from zero ($\beta = -0.60, z = -1.54, p = 0.12$).

Figure 2b indicates that participants preferred a doctor to collect their health information. In Figure 2c, we show the distribution of the comfort and preference ratings of the individuals (across the different categories of health history). This reveals that the preference ratings were strongly bimodally distributed. Most responses clearly preferred a doctor (~ -5) or a virtual assistant ($\sim +5$). This graph also shows that preference was not associated with varying comfort levels. Almost all the comfort levels provided were higher than 4 (on a scale with a maximum of 5).

Discussion Experiment 2

The participants in Experiment 2 reported that they were largely comfortable providing health history data. However, as in Experiment 1, they were somewhat less comfortable providing *Social* history information. Participants preferred a doctor collecting health history, except for the *Social* category. However, even for this least comfortable topic (*Social*), respondents did not show an outspoken preference for the virtual assistant. On average, participants did not prefer either for this topic. When assessing the distributions of the responses, we found that the preference ratings were bimodally distributed. This indicates that most people strongly preferred one or the other actor for each health history aspect. Hence, some people strongly preferred virtual assis-

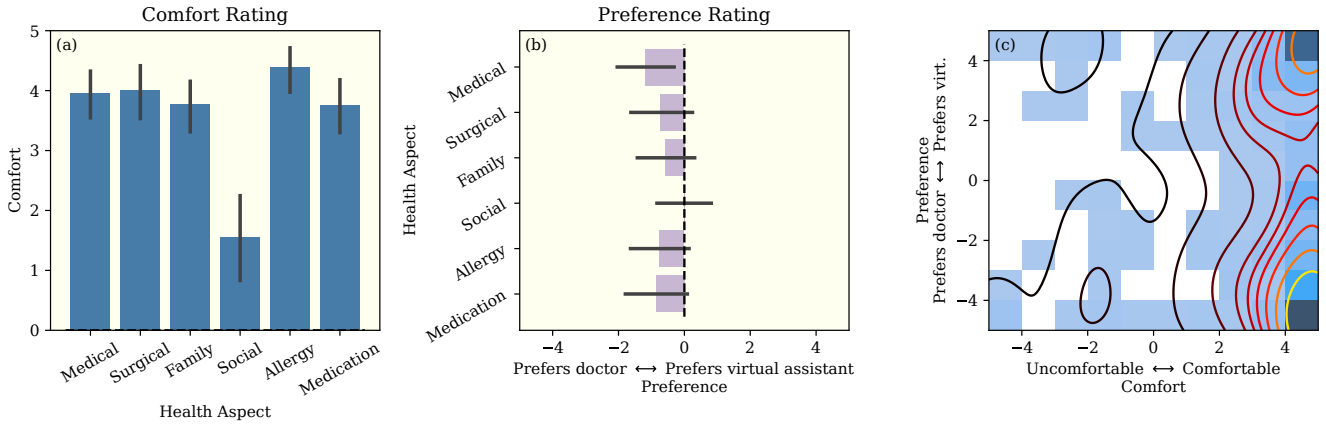

Figure 2: Results of Experiment 2. (a) The average comfort levels for each of the six health history aspects. (b) Preferences for a doctor (negative values) or a virtual assistant (positive values). (c) Heatmap of the reported comfort levels and preferences across the six health aspects. Darker cells indicate a larger proportion of responses. We overlaid a 2D kernel density estimation onto the grid.

tants for some health categories. However, this preference was not associated with varying levels of comfort.

Experiment 3

Experiment 2 assessed whether people preferred their health history to be collected by a doctor or a virtual assistant. This experiment did not reveal a clear preference for a virtual assistant, even for the health aspects respondents were least comfortable discussing. However, this does not necessarily indicate whether people would differ in the amount of disclosure when interviewed by a virtual assistant. Therefore, in Experiment 3, we explicitly asked people how likely they would be to be completely truthful when interviewed by a virtual assistant or a doctor.

Methods Experiment 3

In Experiment 3, we changed the questions presented to the participants. For each category of health history, we asked participants to rate if they would be more comfortable with this information being collected by a human nurse or a virtual assistant. We also asked how complete and truthful they would respond when medical staff collect information. Finally, we asked the same question as if a virtual assistant had collected the information. We recruited undergraduate students at the University of Cincinnati who participated for course credit.

We also split one of the categories of health history information in this experiment into two. In Experiments 1 and 2, we found that the *Social* category was associated with the lowest levels of comfort, truthfulness, and preference for a doctor collecting information. We reasoned that this might be caused by mentioning 'sexual behavior.' Therefore, we created a separate category for sexual behavior to isolate this from other types of health history. This also allowed us to provide a more detailed description of the information that would be collected under this heading. We used the following seven categories and descriptions thereof.

1. **Medical History:** Your medical history includes information about previous diagnoses, past medical interventions, immunization status, dates and results of tests or health screenings, mental health history, and any major illnesses (chronic or acute) you've experienced.
2. **Surgical History:** Your surgical history encompasses details about previous surgical interventions, cesarean section (c-section) deliveries, and instances when you were under general anesthesia.
3. **Family Medical History:** Family medical history refers to diagnoses of medical conditions (e.g., diabetes, cancer, high blood pressure) or mental health conditions (e.g., depression, anxiety, schizophrenia) in your blood relatives.
4. **Social History:** Your social history includes information about your occupation, travel habits, exercise routine, dietary preferences, use of recreational or non-prescription drugs, alternative medicines, any history of prison sentences, and whether firearms are in your household.
5. **Sexual and Reproductive Health History:** Your sexual and reproductive health history involves details about your sexual partners, measures taken to protect against sexually transmitted infections (STIs), any history of STIs, sexual dysfunction, and experiences of sexual trauma or abuse. For individuals assigned female at birth, this also includes information about previous pregnancies/births, miscarriages or abortions, measures taken to prevent unplanned pregnancies, and intentions regarding future pregnancies.
6. **Medication History:** Your medication history includes prescription and non-prescription medications you are currently or have recently taken, previous prescriptions, use of non-prescription drugs like Tylenol or Sudafed, illicit drug use, and any allergies you have to medications.

	Coef.	Std. Err.	z	P> |z|
Intercept	-0.542	0.380	-1.425	0.154
T.1	-1.558	0.281	-5.538	0.000
T.2	-1.726	0.290	-5.956	0.000
T.3	-1.668	0.283	-5.892	0.000
T.4	-1.119	0.279	-4.013	0.000
T.6	-1.333	0.293	-4.555	0.000
T.7	-1.582	0.295	-5.369	0.000
Intrinsic	-0.289	0.480	-0.602	0.547
Group Var	5.924	0.509		

Table 3: Experiment 3, Comfort rating

	Coef.	Std. Err.	z	P> |z|
Intercept	2.952	0.230	12.812	0.000
T.1	0.837	0.169	4.957	0.000
T.2	0.874	0.169	5.164	0.000
T.3	0.719	0.168	4.269	0.000
T.4	0.259	0.168	1.538	0.124
T.6	0.874	0.169	5.179	0.000
T.7	1.356	0.169	8.038	0.000
Truth, VA	-0.264	0.090	-2.935	0.003
Intrinsic	0.173	0.280	0.619	0.536
Group Var	2.098	0.175		

Table 4: Experiment 3, Truthful rating

7. **Allergy History:** Your allergy history encompasses seasonal allergies, food allergies (e.g., tree nut allergy), food sensitivities (e.g., gluten intolerance), any unexplained rashes or reactions, and medication allergies.

We introduced a between-subject variable by varying the context presented to the participants. Half of the participants were told they were visiting a new doctor for the first time, expecting this doctor to become their primary care provider. The health history collected would enable the doctor to provide the best possible care. This condition was designed to elicit the intrinsic motivation of the respondents to provide accurate information, as the information would be used to benefit them directly.

We also introduced an extrinsic motivation variation of the scenario. In this condition, participants were told they were required to see a doctor as part of onboarding for a new job. During their first visit, the doctor's office collects their health history. This information lets the doctor inform the new employer how to insure them at work. It is important that the information provided is accurate and complete, regardless of an illness, injury, or mental health diagnosis. We told participants that they would not be punished or fired for having a (or any) condition, injury, or mental health diagnosis. However, they could be punished at work or fired for providing inaccurate or incomplete information. By introducing this variation, we attempted to understand whether the preference for a human or virtual assistant collecting the health history depended on the context, i.e., the beneficiary of the correct information collection.

Results Experiment 3

The survey was completed by 144 respondents. Their demographics were as follows. Age: 18 (34%), 19 (43%), 20 (16%), 21 (3%). Ethnicity: Asian (8%), Black or African American (5%), Other (18%), White (69%). Gender: Female (69%), Male (24%), Non-binary (5%).

We ran a mixed linear model predicting the comfort rating of the participants as a function of the health history category and the type of motivation. The results of this model are given in Table 3. This analysis did not reveal a main effect of the type of motivation. We also did not find a pattern of interaction effects. Therefore, we will further analyze the comfort levels collapsed across the between-subject variable *motivation*.

Figure 3 presents the results of Experiment 3. Figure 3a shows that, in general, participants would be more comfortable with a doctor collecting health history (negative values). The difference in comfort was the least outspoken for the *Sexual* health category. Participants were significantly more comfortable with a doctor for the other categories (see Table 3).

Figure 3b shows that people indicated they would be more truthful when the doctor collected the medical history. This difference was statistically significant (using a mixed linear model, $(p < 0.01)$, Table 4). However, as seen in Figure 3b, this did not hold for the *Sexual* category. For this category, respondents indicated they would be more likely to be truthful if a virtual assistant collected their data. However, this difference was not statically significant ($\beta = 0.50, p = 0.12$)

Next, we analyze the relationship between the level of comfort and the intention of responding truthfully. For each participant and health aspect, we calculated the difference in the intention to respond truthfully if a doctor collected the health history and if the virtual assistant collected the health history. We visualized this difference as a function of the comfort level (see Figure 3c). This shows that when people are more comfortable with a virtual assistant (doctor), they are more inclined to be truthful if it (they) collects their health history.

Discussion Experiment 3

The comfort ratings in Experiment 3 were the lowest for the *Sexual* category (Figure 3a, Table 3). However, even for this category, participants tended to be more comfortable with a doctor collecting their health history. Moreover, participants reported they would be slightly more truthful when a doctor collects their health history (Table 4). Participants were more honest with a virtual assistant if they felt more comfortable with it, and vice versa. (Figure 3c). Therefore, in Experiment 3, we found some (limited) evidence that a virtual assistant might reduce nondisclosure when participants are more comfortable with it. For the *Sexual* category, participants were slightly more likely to be truthful with a virtual assistant. However, this difference was not statistically significant.

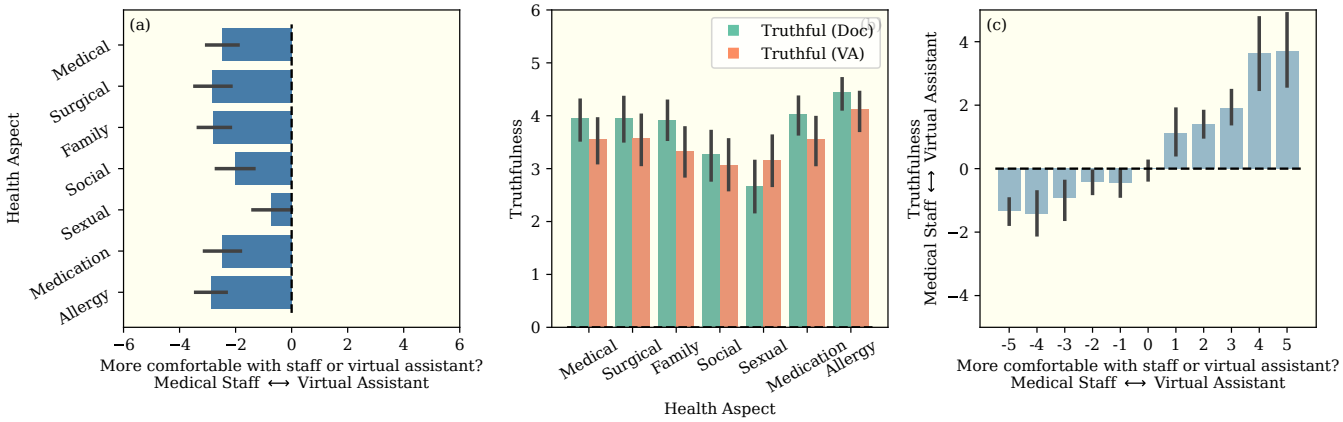

Figure 3: Results of Experiment 3. (a) Average comfort rating (negative values indicate that respondents were more comfortable with a doctor collecting the health history). (b) The average level of truthfulness when the health history is collected by a doctor (Doc) or virtual assistant (VA). (c) Difference in truthfulness as a function of the comfort rating across the health history categories.

Experiment 4

In Experiments 2 and 3, people indicated they would be more comfortable with a doctor collecting their health history. They also responded that they would be more truthful when interviewed by a doctor (except for the *Sexual* category). In the scenarios presented so far, the virtual assistant and the doctor are both presented as entirely passive: they are only said to collect information from the patient. However, besides collecting information, virtual assistants are also envisioned as devices that could provide some healthcare interventions (Laranjo et al. 2018). The failure to find clear benefits for virtual assistants while collecting healthcare information led us to ask how patients would respond to virtual assistants making intervention decisions. Hidalgo et al. (2021), Vanderelst and Willems (2020), and Vanderelst et al. (2023) assessed how people view robots making medical decisions (among others). Here, we adopt their methodology to evaluate respondents' perception of virtual assistants' actions after collecting their health history.

Methods Experiment 4

In Experiment 4, participants were presented with a scenario involving Robert, an elderly patient who visits a new doctor for the first time. The scenario described Robert's health information, which a nurse or a virtual assistant collected. Participants were randomly assigned to a scenario that involved a nurse or virtual assistant. The participants were also presented with one of the following three outcomes (or actions taken by the nurse or the virtual assistant). In the first variation of the scenario, the respondents were told that the actor (nurse or virtual assistant) believed that Robert could be in an abusive relationship based on his health history and physical symptoms and decided to seek external intervention for Robert. The second variation of the scenario involved the actor finding that Robert's health history suggests a contagious illness that could threaten public health and decides to report it. Finally, in the third version, participants were told that the actor believes that Robert may be suicidal based on his health history and decides to seek external intervention for Robert. We told participants that several research groups are developing AI-based virtual assistants (like Google Assistant, Apple's Siri, or Amazon's Alexa) to help care for patients. We also mentioned that, among other things, these virtual assistants might help collect medical and health history information.

Participants rated the scenario on 14 scales using a 7-point labeled Likert scale. These 14 scales included three scales used by Vanderelst et al. (2023) (scales 0-2, Table 5) and 11 scales used by Hidalgo et al. (2021) to assess the differences in people's judgments of humans and machines. Table 5 lists the prompts associated with each scale.

After completing the 14 scales, participants were asked to pick four words from 20 that best described the scenario they just read (listed in Table 6). Hidalgo et al. (2021) presented their participants with these words to measure the moral dimensions associated with each scenario.

After rating the scenario with the nurse or virtual assistant as an actor, the participants were presented with the same scenario but with the other actor. Again, they were asked to rate the scenario on the 14 scales presented in Table 5 and select words from those listed in Table 6.

We recruited 300 subjects using Prolific and compensated them $2.00 for participating. We requested a representative sample of the United States. After participants completed the main ratings, they were invited to provide several demographics.

Results Experiment 4

Aggregated Scales For Experiment 4, we collected the participants' ratings on 14 7-point Likert scales. We evaluated whether these scales could be collapsed into fewer scales by assessing their pairwise correlations to facilitate data analysis.

Scale nr	Scale prompt	Aggregated scale
0	Was the actor's action acceptable?	Ethical
1	Was the actor's action ethical?	Ethical
2	Was the actor's action the right thing to do?	Ethical
3	Was the actor's action harmful?	Ethical
4	Would you hire this actor for a similar position?	Ethical
5	Was the actor's action intentional?	Intentional
6	Do you like the actor?	Ethical
7	How morally wrong or right was the actor's action?	Ethical
8	Do you agree that the actor should be promoted to a position with more responsibilities?	Ethical
9	Do you agree that the actor should be replaced by the other actor?	Replace other
10	Do you agree that the actor should be replaced by another actor?	Replace same
11	Do you think the actor is responsible for the action?	Intentional
12	Do you think the actor is responsible for the outcome?	Intentional
13	If you were in a similar position as the actor, would you have done the same?	Ethical

Table 5: This table provides an overview of the 14 scales used by participants to rate the presented scenarios. The first column denotes the index referenced in Figure 4 for each scale. The second column presents the prompts presented to participants. The third column displays the aggregated scales derived from multidimensional scaling of the original scales.

Words presented to participants			
Harmful	Discriminatory	Devoted	Respectful
Violent	Fair	Loyal	Indecent
Caring	Impartial	Disobedient	Obscene
Protective	Disloyal	Defiant	Decent
Unjust	Traitor	Lawful	Virtuous

Table 6: Words presented to participants. They selected four words that best described the scenario.

We calculated the pairwise polychoric correlation coefficients between the 14 scales for both experiments. Next, we converted the resulting correlation matrix into a distance matrix by taking the absolute value and subtracting the result from 1. The result of this operation was visualized using two-dimensional metric scaling, as shown in Figure 4.

Based on Figure 4, we aggregated scales 0, 1, 2, 3, 4, 6, 7, 8, and 13 into a single scale by averaging across them (after inverting scale 3). In the remainder of the article, we refer to this aggregated scale as the Ethical aggregated scale. We use this denominator for the scale as its comprising scales measure the harmfulness and ethicality of the action. Scales 5, 11, and 12 in Table 5 were also aggregated into a single scale we will refer to as the Intentional scale. Scales 9 and 10 were analyzed separately.

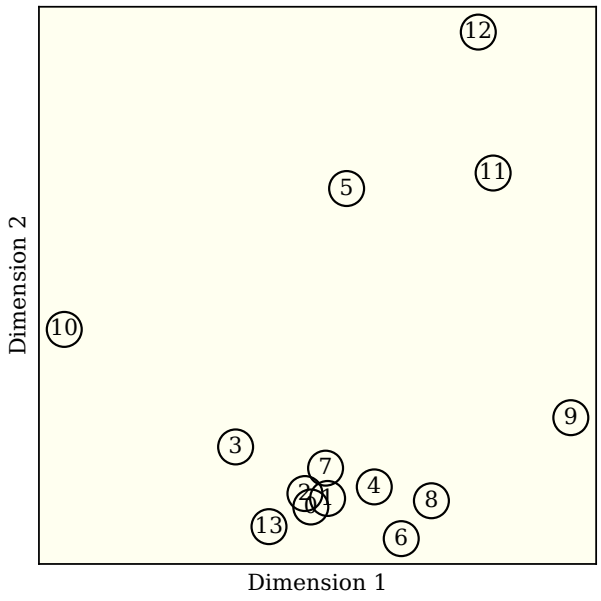

Figure 4: Results of two-dimensional metric multidimensional scaling of the 14 scales used in Experiment 3. The numbers of the scales correspond to those listed in Table 5.

Analysis The demographics are visualized and compared with US census data in Figure 5. We statistically analyzed the results of Experiment 3 using a mixed linear model with actor and action as categorical predictors. We did not include interaction effects in the model. We use a mixed linear model with the respondents as a random variable. The dependent variable was the rating on each of the four aggregated scales. The effect of the actor is the focus of this article. Therefore, we report on this here. In Figure 6, we visualize the data collapsed across this variable.

Table 7 lists the results of the mixed linear model. This reveals that the effect of the actor (nurse versus virtual assistant) was significant ($p < 0.01$). The same actions or decisions taken by the virtual assistant were deemed less ethical than if performed by a nurse. This can also be seen in Figure 6. The size of the effect was about 1 on a 7-point scale.

The action was also seen to be more intentional when performed by a nurse than if performed by the virtual assistant ($p < 0.01$). The respondents had a clear preference for the nurse. They clearly tended to want to replace the virtual assistant with a nurse rather than the other way around ($p < 0.01$). Finally, respondents expressed wanting to re-

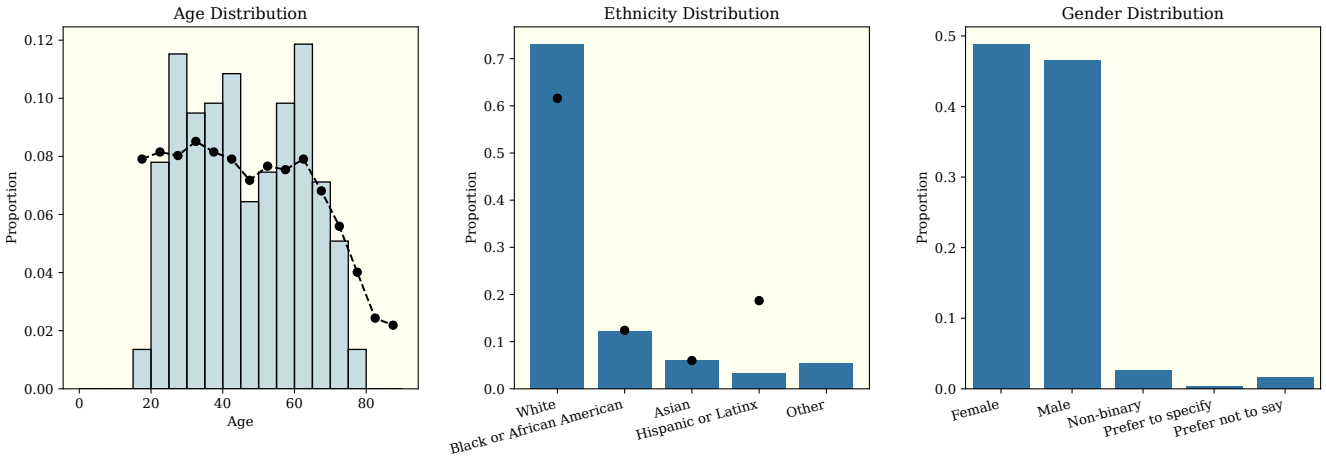

Figure 5: Vizualization of the self-reported demographics in Experiment 4. Age was obtained by subtracting the reported birth year from 2022 (the year the study was conducted). We overlaid the age distribution for the US (source: U.S. Census Bureau, 2022). For ethnicity, we plotted the proportions as observed in the 2020 Census Redistricting data (which does not include Puerto Rico).

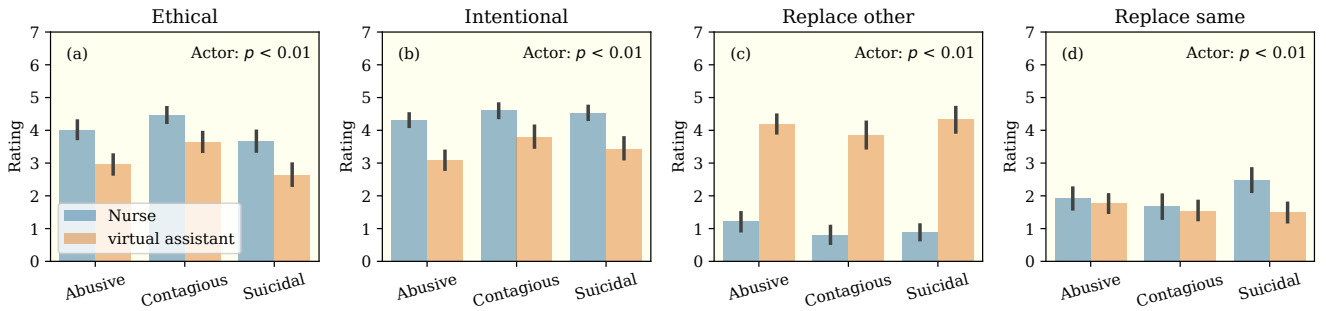

Figure 6: Results of Experiment 4. (a) Average rating on the aggregated *Ethical* scale. (b) Average rating on the aggregated *Intentional* scale. (c) Average rating on the *Replace Other* scale. (d) Average rating on the *Replace Same* scale. The components of the scales are listed in Table 5.

	Coef.	Std. Err.	z	P> |z|
Intercept	3.974	0.135	29.361	0.000
Actor	-0.975	0.067	-14.501	0.000
Contagious	0.565	0.191	2.955	0.003
Suicidal	-0.339	0.188	-1.805	0.071
Group Var	1.482	0.256		

Table 7: Mixed Linear Model Regression Results. Experiment 4, scale: Ethical

place the virtual assistant with another virtual assistant, more than a nurse with another nurse ($p < 0.01$).

Figure 7 shows the results of the word associations selected by the participants for each actor and action. The two most selected words were *protective* and *caring*. However, these were selected more often for a nurse than for a virtual assistant. This could indicate that the nurse was perceived to be more caring and protective of Robert. Interestingly, the virtual assistant was more associated with the words *fair* and *impartial*.

Discussion Experiment 4

Experiment 4 revealed that the same action, when performed by a virtual assistant, was deemed slightly less ethical than when performed by a nurse. This indicates that the respondents judged the virtual assistant's actions more harshly, although they thought its actions were less intentional than those of the nurse. This was also reflected in the word association data. The nurse was perceived as more caring, while the virtual assistant was deemed more fair and impartial.

The results mimicked those of the other experiments in that the nurse was more liked. Participants were more keen on replacing the virtual assistant with the nurse than the other way around. If they wanted to replace the nurse, they had a slight preference for replacing them with another nurse instead of a virtual assistant.

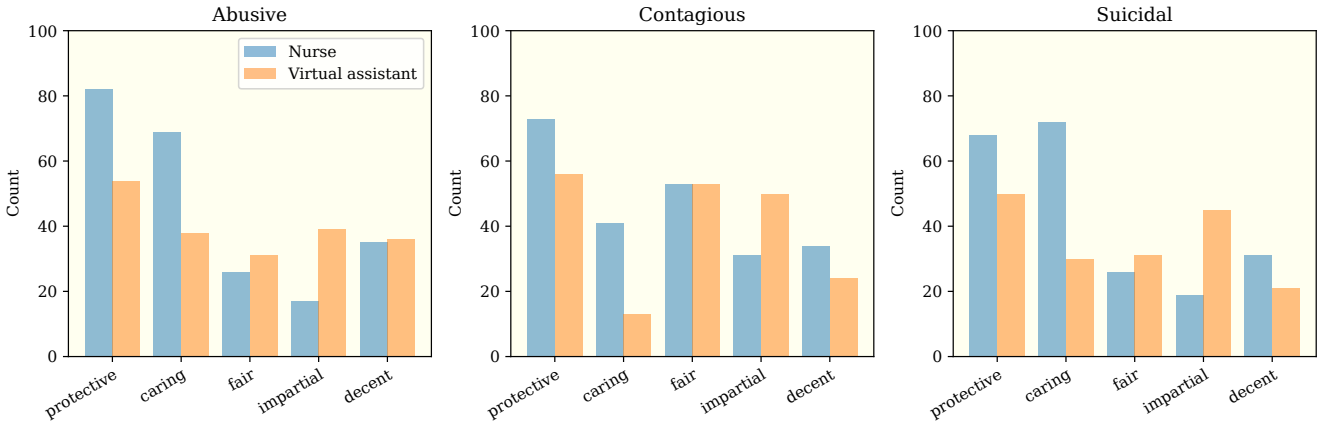

Figure 7: Results of the word associations collected in Experiment 4. For clarity, we only visualize the five most commonly selected words.

General Discussion

Lucas et al. (2014) suggested that virtual assistants might reduce the psychological barriers causing non-disclosure by increasing a patient's sense of anonymity. Moreover, unlike classic computer-administered assessments, virtual assistants could result in rapport with the patients. As such, Lucas et al. (2014) claimed that virtual assistant interviewers could offer the best of two worlds. If virtual assistants reduce non-disclosure, their use is ethically imperative as non-disclosure can have severe consequences for a patient's health (Levy et al. 2018).

We presented data from four online experiments to assess whether virtual assistants could decrease non-disclosure in different areas of a person's medical history. Overall, we found little evidence of the usefulness of virtual assistants in this regard. Experiment 1 confirmed that the various health history aspects we considered elicit different levels of disclosure comfort and likelihood of responding completely and truthfully. Therefore, even though we asked people to imagine a scenario, they perceived the various aspects of their health history as differently sensitive or stigmatizing (Figure 1). In particular, *Social* health aspects were associated with lowered comfort and intention to tell the truth. However, in Experiment 2, we did not find evidence that this led to a preference for health history to be collected by a virtual assistant. Participants indicated they preferred to have their health history collected by a doctor. Even for the *Social* health aspects, people did not prefer a virtual assistant but were virtually neutral (Figure 2a). Visualizing individual responses, we could not find an association between the level of reported comfort and a person's preference for the virtual assistant or a doctor (Figure 2b). This analysis revealed that, in these data, participants' preferences were bimodally distributed: some people strongly preferred a doctor, while others strongly preferred a virtual assistant. However, this preference was not associated with a person's reported comfort level.

In Experiment 3, we did not ask respondents about their preferences. Instead, we asked how inclined they would be to respond truthfully to a doctor and a virtual assistant (their intention to disclose their health history fully). Respondents indicated they were *less* inclined to respond truthfully when interviewed by a virtual assistant (Figure 3a Table 4), except for the *Sexual* health category. For this health history aspect, we found that people would be more likely to be truthful if interviewed by a virtual assistant. However, this difference was small and was not statistically significant. Experiment 3 confirmed that, in agreement with the existing literature (as reviewed by Lucas et al. 2014), people are more inclined to tell the truth if they are more comfortable with the interviewer. If people were more comfortable being interviewed by a doctor (virtual assistant), they indicated they would be more inclined to tell the truth when interviewed by this actor.

In Experiments 1, 2, and 3, the actors were portrayed as purely passive collectors of health history information. In Experiment 4, we assessed the perception of healthcare interventions by virtual assistants. We asked people to assess how ethical and likable a nurse or virtual assistant was when making a potentially ethically ambivalent decision. In particular, we presented scenarios where a nurse or a virtual assistant made an active healthcare intervention. Vanderelst and Willems (2020), Vanderelst et al. (2023), and Hidalgo et al. (2021) used this methodology to assess differences in the way people judge other humans and artificial agents.

Experiment 4 showed that people judged the virtual assistant's actions as somewhat less ethical, even though they saw it as less responsible for its actions. Moreover, people strongly preferred that nurses replace virtual assistants. These results indicate that when a virtual assistant makes a judgment with ethical implications, it would be judged more harshly than a human (and less caring). These data suggest that a virtual assistant might find it more challenging to establish and maintain rapport with a patient than a human after a healthcare intervention.

Several studies have investigated how virtual assistants (in healthcare) could be designed to elicit more disclosure (See Curtis et al. 2021, for a review). These studies found, for ex-

ample, that virtual assistants who are likable or disclose information about themselves are more likely to be trusted by interviewees. However, few studies have assessed whether virtual assistants elicit more disclosure than human interviewees. One exception is the work of Lucas et al. (2014).

Lucas et al. (2014) had participants perform mock interviews in which a virtual assistant asked about clinical symptoms. In one condition, people were led to believe the virtual assistant was autonomous. In another condition, participants thought they interacted with a virtual assistant controlled by a human. Lucas et al. (2014) found that respondents were less fearful of self-disclosure and more willing to disclose.

The results of Lucas et al. (2014) seem to contradict the current results, which find little evidence for the benefit of virtual assistants. However, on closer inspection, the results of Lucas et al. (2014) might align more than initially thought. Indeed, the effect sizes reported by Lucas et al. (2014) are small. People being interviewed by a (said to be) fully autonomous virtual assistant were more fearful than those interviewed by a virtual assistant (said to be) controlled by a human. However, the difference was small: $\sim$2 (or $\sim 4\%$) on a scale ranging from 12 to 60. Likewise, the rated willingness to disclose, although statistically significant, differed only by about 0.54 on a scale from -3 to +3. The statistical significance of these small effect sizes is due to a sample size of about 240 respondents across the two conditions.

The current methodology, asking people to imagine being interviewed by a virtual assistant, differs from the mock interviews conducted by Lucas et al. (2014). On the one hand, the mock interviews of Lucas et al. (2014) seem more ecologically valid as they more closely mimicked the situation under study. On the other hand, they did not directly compare a virtual assistant with a human interviewer. In Lucas et al. (2014), the human interviewer condition was still a virtual agent, albeit one that was said to be controlled by a human. Being interviewed by a human controlling a virtual assistant might have been perceived as awkward by the participants. They were effectively asked to divulge information to a human interviewer they could not see. This might explain why they are less comfortable and less inclined to disclose information when interviewed by a human. Therefore, even though they report minor effects, these could exaggerate the benefits of a virtual assistant as they compare it with a particularly adverse alternative.

We summarize the current results and those of Lucas et al. (2014) as indicating that the ability of virtual assistants to reduce non-disclosure is limited. With hindsight, the finding that virtual assistants have a limited impact on the patients' disclosure is perhaps easy to explain. Non-disclosure is primarily driven by a fear of being judged by the healthcare provider. A virtual assistant collecting health history would still relay that information to a human healthcare provider. Indeed, if the virtual assistant would not disclose information, it would not benefit the patient. Therefore, patients may reason that providing information to a virtual assistant is just a roundabout way of telling a doctor or nurse. The limited impact on disclosure combined with their potential ethical issues pointed out by other authors (Luxton 2020; Del Valle, Llorca Albareda, and Rueda 2024; Luxton 2014) suggests that, while virtual assistants could have practical advantages, their net ethical impact is likely to be negative.

References

Abrams, Z. 2024. Addressing Equity and Ethics in Artificial Intelligence. *Monitor on Psychology*, 55(3).

Coeckelbergh, M. 2020. Challenges for Policymakers. In *AI Ethics*, 0. The MIT Press. ISBN 978-0-262-35706-7.

Curtis, R. G.; Bartel, B.; Ferguson, T.; Blake, H. T.; Northcott, C.; Virgara, R.; and Maher, C. A. 2021. Improving User Experience of Virtual Health Assistants: Scoping Review. *Journal of Medical Internet Research*, 23(12): e31737.

Del Valle, J. I.; Llorca Albareda, J.; and Rueda, J. 2024. Ethics of Virtual Assistants. In *Ethics of Artificial Intelligence*, 87–107. Springer.

Hidalgo, C. A.; Orghian, D.; Canals, J. A.; De Almeida, F.; and Martin, N. 2021. *How humans judge machines*. MIT Press.

Laranjo, L.; Dunn, A. G.; Tong, H. L.; Kocaballi, A. B.; Chen, J.; Bashir, R.; Surian, D.; Gallego, B.; Magrabi, F.; Lau, A. Y. S.; and Coiera, E. 2018. Conversational Agents in Healthcare: A Systematic Review. *Journal of the American Medical Informatics Association: JAMIA*, 25(9): 1248–1258.

Levy, A. G.; Scherer, A. M.; Zikmund-Fisher, B. J.; Larkin, K.; Barnes, G. D.; and Fagerlin, A. 2018. Prevalence of and Factors Associated With Patient Nondisclosure of Medically Relevant Information to Clinicians. *JAMA Network Open*, 1(7): e185293.

Lucas, G. M.; Gratch, J.; King, A.; and Morency, L.-P. 2014. It's Only a Computer: Virtual Humans Increase Willingness to Disclose. *Computers in Human Behavior*, 37: 94–100.

Luxton, D. D. 2014. Recommendations for the Ethical Use and Design of Artificial Intelligent Care Providers. *Artificial Intelligence in Medicine*, 62(1): 1–10.

Luxton, D. D. 2020. Ethical implications of conversational agents in global public health. *Bull. World Health Organ.*, 98(4): 285–287.

Luxton, D. D.; and Watson, E. 2023. Psychological and Psychosocial Consequences of Super Disruptive A.I.: Public Health Implications and Recommendations. In *Proceedings of the Stanford Existential Risk Conference*.

Milne-Ives, M.; de Cock, C.; Lim, E.; Shehadeh, M. H.; de Pennington, N.; Mole, G.; Normando, E.; and Meinert, E. 2020. The Effectiveness of Artificial Intelligence Conversational Agents in Health Care: Systematic Review. *J. Med. Internet Res.*, 22(10): e20346.

Piñeiro Martín, A.; García Mateo, C.; Docío Fernández, L.; and del Carmen López Pérez, M. 2022. Ethics Guidelines for the Development of Virtual Assistants for e-Health . In *Proc. IberSPEECH 2022*, 121–125.

Vanderelst, D.; Jorgenson, C.; Ozkes, A. I.; and Willems, J. 2023. Are Robots to be Created in Our Own Image? Testing the Ethical Equivalence of Robots and Humans. *International Journal of Social Robotics*, 15(1): 85–99.

Vanderelst, D.; and Willems, J. 2020. Can we agree on what robots should be allowed to do? An exercise in rule selection for ethical care robots. *International Journal of Social Robotics*, 12: 1093–1102.