TESIS DE LA UNIVERSIDAD DE ZARAGOZA
2025
329

Guillermo Calahorra Candao

Understanding purchase intention in voice commerce: the effects of voixe type through anthropomorphism, trust, safety, and emotional response

Director/es
Martín De Hoyos, María José

http://zaguan.unizar.es/collection/Tesis

ISSN 2254-7606

Prensas de la Universidad
Universidad Zaragoza

Universidad de Zaragoza
Servicio de Publicaciones

ISSN 2254-7606



Universidad Zaragoza
1542

Tesis Doctoral

UNDERSTANDING PURCHASE INTENTION IN VOICE COMMERCE: THE EFFECTS OF VOIXE TYPE THROUGH ANTHROPOMORPHISM, TRUST, SAFETY, AND EMOTIONAL RESPONSE

Autor

Guillermo Calahorra Candao

Director/es

Martín De Hoyos, María José

UNIVERSIDAD DE ZARAGOZA
Escuela de Doctorado

Programa de Doctorado en Economía y Gestión de las Organizaciones

2025

Repositorio de la Universidad de Zaragoza – Zaguan http://zaguan.unizar.es

TESIS DOCTORAL

UNDERSTANDING PURCHASE INTENTION IN VOICE COMMERCE:

THE EFFECTS OF VOICE TYPE THROUGH ANTHROPOMORPHISM, TRUST, SAFETY, AND EMOTIONAL RESPONSE

Autor:

Guillermo Calahorra Candao

Directora:

María José Martín de Hoyos

Facultad de Economía y Empresa

2025

Repositorio de la Universidad de Zaragoza – Zaguan http://zaguan.unizar.es

“I’m nothing more than recycled stardust and borrowed energy.”

— Kesha, “Spaceship”

AGRADECIMIENTOS

Durante estos años, solía *bromear* sobre mi estado físico y mental al finalizar la tesis. Compaginar un trabajo a jornada completa, independiente de la tesis, junto con otras responsabilidades paralelas, ha sido un desafío que, en más de una ocasión, me ha traído algún que otro quebradero de cabeza. Sin embargo, hay personas que han hecho que este viaje haya sido mucho más llevadero y a quienes me gustaría dar las gracias.

Quiero empezar dando las gracias a María José, por creer en lo que estaba haciendo, por acompañarme y guiarme en el camino de la investigación académica mucho antes de empezar la tesis doctoral, y por impulsarme a repensar y mejorar cada idea. No puedo pensar en un mejor referente y mentora para toda esta aventura, y me siento profundamente afortunado por haber recorrido este camino a su lado.

Me gustaría agradecer a los miembros del departamento de Dirección de Marketing e Investigación de Mercados la colaboración prestada durante estos años, especialmente al grupo de investigación GENERES. Gracias a Carol, por aparecer siempre que la he necesitado y por su constante disposición a ayudar; y a Sara, con quien compartí los inicios de la tesis y que siempre ha tenido una palabra agradable para mí.

A mis padres, por darme la libertad de elegir mi propio camino y enseñarme el valor del trabajo, la constancia y la fuerza de voluntad.

A Arturo, por estar a mi lado en todo momento, en los días buenos y en los no tan buenos, celebrando conmigo los logros y sosteniéndome en los momentos difíciles. Gracias por hacerme reír y darme siempre el plato de comida bonito.

A mis amigos, por todo su afecto, compañía y apoyo en este proceso. En especial, a Dani, por hacer que la distancia no pusiese barreras a nuestra amistad; a Javi y Marco, por hacer que cada encuentro sea especial; a Sergio y Jorge S., por sus consejos, tanto personales como académicos, y hacerme sentir *en casa*; a Jorge O. y Javi, por su cariño desde el minuto en el que les molesté bailando; y a Clara, porque con ella sí ha sido *a best friend race*.

Al personal y compañeros del hospital de día del Hospital Clínico de Zaragoza, que, de manera inesperada, terminaron siendo testigos y espectadores de muchas de las intervenciones, charlas y ponencias que el destino hizo coincidir con mis tratamientos. En especial, a Mari Carmen, cuyas constantes palabras de apoyo y cariño consiguieron que ese “chute” fuese también de ánimos y esperanza.

Y por último, a quienes por distintos motivos ya no están a mi lado, pero en su momento lo estuvieron y sé que estarían orgullosos de lo que he conseguido. Edu, allá donde estés, espero que veas volar cometas por el cielo.

🛸
✨
🐄

INDEX OF CONTENTS

CHAPTER 0: Introducción 2

Referencias 15

CHAPTER 1: Study 1 - Acceptance and Use of Voice Assistants 19

1.1. Introduction 20

1.2. Voice shopping acceptance literature review 20

1.3. Proposed model and hypothesis development 28

1.3.1. Perceived Usefulness and Perceived Ease of Use 29

1.3.2. Perceived Safety 29

1.3.3. Enjoyment 29

1.3.4. Performance Expectancy 30

1.3.5. Social Norms 30

1.4. Methodology 34

1.5. Results 35

1.5.1. Testing the measurement model 35

1.5.2. Testing the structural model 38

1.6. Conclusions and discussion 40

1.6.1. Implications for theory 42

1.6.2. Implications for practice 43

1.6.3. Limitations and future research 43

1.7. References 45

1.8. Appendix 55

CHAPTER 2: Study 2 - Perceived Safety and the Humanization of Voice Assistants 57

2.1. Introduction 58

2.2. Literature review 60

2.2.1. Anthropomorphism 60

2.2.2. Safety 64

2.2.3. Voice shopping 66
2.3. Conceptual framework and hypotheses development 68
2.3.1. Anthropomorphic variables 68
2.3.2. Safety perception 70
2.4. Methodology 72
2.5. Results 74
2.5.1. Testing the measurement model 74
2.5.2. Testing the structural model 76
2.6. Conclusions and discussion 78
2.6.1. Implications for theory 80
2.6.2. Implications for practice 81
2.6.3. Limitations and future research 82
2.7. References 84
2.8. Appendix 98

CHAPTER 3: Study 3 - Trusting Virtual Assistants: The Power of Emotional Design 100

3.1. Introduction 101
3.2. Theoretical framework 104
3.2.1. Emotional response to basic emotions 104
3.2.2. Human-like voice 105
3.2.3. Trust in VAs 106
3.2.4. The moderating effect of voice type 108
3.2.5. Impact of Message Valence on Emotional Response and Trust 109
3.3. Methodology 110
3.3.1. Study design 110
3.3.2. Sample 112
3.3.3. Measuring variables 114
3.4. Analyses and results 117
3.4.1. Analysis I: Model estimation 117
3.4.2. Analysis II: Moderating effect of voice type 120
3.4.3. Analysis III: Automatic emotion detection 121
3.4.4. Analysis IV: Moderating effect of message valence 122

3.5. Conclusions and discussion 124
3.5.1. Key findings 124
3.5.2. Implications for theory 126
3.5.3. Implications for practice 128
3.5.4. Limitations and future research 130
3.6. References 132
3.7. Appendix 142

CHAPTER 4: Study 4 - The Impact of Voice Type on Purchase Intention in Voice Shopping 143
4.1. Introduction 144
4.2. Type of Voice in Voice Shopping: Theoretical Framework and Hypothesis Development 150
4.2.1. Direct effect 150
4.2.2. Mediated effects 151
4.2.3. The Moderating Role of Voice Type 156
4.3. Research Methodology 159
4.3.1. Voice characteristics analysis 159
4.3.2. Methodological approach for moderation models 160
4.3.3. Measuring variables 162
4.4. Analyses and results 163
4.4.1. Model 1 164
4.4.2. Model 2 165
4.4.3. Model 3 166
4.4.4. Model 4 168
4.4.5. Model 5 170
4.4.6. Model 6 171
4.4.7. Model 7 172
4.5. Conclusions and discussion 175
4.5.1. Implications for theory 179
4.5.2. Implications for practice 180
4.5.3. Limitations and future research 181
4.6. Conclusions 182

4.7. References 184
4.8. Appendix 190

CHAPTER 5: Study 5 - Conclusiones 192
5.1. Conclusiones 193
5.2. Implicaciones teóricas 196
5.3. Implicaciones empresariales 198
5.4. Limitaciones y futuras líneas 203
5.5. Referencias 205

INDEX OF TABLES

CHAPTER 0

Tabla 0.1. Modelos conceptuales 14

CHAPTER 1

Table 1.1. Voice shopping acceptance studies 23
Table 1.2. Measurement model results 36
Table 1.3 Correlations among factors 37
Table 1.4. Heterotrait-monotrait ratio (HTMT) 37
Table 1.5. Proposed hypotheses 39
Table 1.6. Indirect effects 39
Table 1.7. Scales 54

CHAPTER 2

Table 2.1. Age, gender and VA distribution 74
Table 2.2. Measurement model results 75
Table 2.3. Correlations among factors 76
Table 2.4. Heterotrait-monotrait ratio (HTMT) 76
Table 2.5. Proposed hypotheses 77
Table 2.6. Bootstrap multigroup analysis results 78
Table 2.7. Scales 98

CHAPTER 3

Table 3.1. Measurement model results 117
Table 3.2. Correlations among factors 118
Table 3.3. Heterotrait-monotrait ratio (HTMT) 118
Table 3.4. Effect size f^2 119
Table 3.5. Predictive index Q^2 118
Table 3.6. Multigroup analysis 120
Table 3.7. Differences among positive and negative messages 123
Table 3.8. Scales 142

CHAPTER 4

Table 4.1. Measurement model results 164
Tabla 4.2. Voice effects on purchase intention 174
Table 4.3. Scale 190

CHAPTER 5

Tabla 5.1 Resumen de las principales conclusiones 202

INDEX OF FIGURES

CHAPTER 1

Figure 1.1. Proposed model 34
Figure 1.2. Estimated model 38

CHAPTER 2

Figure 2.1. Proposed model 72
Figure 2.2. Estimated model 77

CHAPTER 3

Figure 3.1. Conceptual Model and Hypothesis 108
Figure 3.2. Spectral distribution of synthetic voices 112
Figure 3.3. Data collection procedure 114
Figure 3.4. Circumplex model of affect 116
Figure 3.5. Estimated model 119
Figure 3.6. Classification of automatic emotion detection 122

CHAPTER 4

Figure 4.1. Model 1 151
Figure 4.2. Model 2 153
Figure 4.3. Model 3 154
Figure 4.4. Model 4 156
Figure 4.5. Model 5 157
Figure 4.6. Model 6 158
Figure 4.7. Model 7 158
Figure 4.8. Voice characteristics 160
Figure 4.9. Model 1 results 165
Figure 4.10. Model 2 results 166
Figure 4.11. Model 3 results 168
Figure 4.12. Model 4 results 170
Figure 4.13. Visualization of interaction effect 171
Figure 4.14. Visualization of interaction effect 174

RESUMEN

A medida que los asistentes de voz (AV) se integran cada vez más en nuestras rutinas diarias, la forma en que compramos está cambiando: pasamos de teclear y desplazarnos por pantallas a hablar y escuchar. Sin embargo, a pesar de su creciente presencia en los hogares, las compras por voz aún enfrentan barreras relacionadas con la confianza del usuario y su adopción. Esta tesis investiga los mecanismos psicológicos, emocionales y perceptivos que influyen en cómo los consumidores interactúan con los AV al tomar decisiones de compra.

La investigación adopta un enfoque centrado en las personas, explorando cómo responden emocionalmente los usuarios a las voces sintéticas en contextos comerciales. Basándose en modelos de aceptación tecnológica, se argumenta que la funcionalidad por sí sola no basta. Para que las compras por voz tengan éxito, los usuarios deben sentir que el sistema funciona correctamente, respeta su privacidad y establece una conexión emocional con ellos. Conceptos como la seguridad percibida, el disfrute, la confianza y la respuesta emocional emergen como factores clave.

Uno de los focos principales del estudio es la antropomorfización. Características como una voz natural, la amabilidad o la sensación de presencia social contribuyen significativamente a generar confianza y sensación de seguridad. Los asistentes percibidos como más humanos provocan percepciones más positivas y una mayor disposición a comprar mediante la voz. La seguridad no se presenta sólo como un aspecto técnico, sino también como un fenómeno psicosocial. La respuesta emocional ocupa un lugar central en esta tesis, influyendo notablemente en cómo los usuarios perciben a los asistentes de voz y si confían en ellos. Las emociones positivas generadas por el asistente refuerzan la percepción de humanidad e incrementan la confianza, mientras que las emociones negativas reducen ambos aspectos. Además, el tono emocional del mensaje también importa: la valencia del contenido modera esta relación, potenciando el efecto de la respuesta emocional sobre la percepción de humanidad cuando el contenido es positivo.

Un estudio experimental con 900 participantes simuló interacciones de compra mediante voz. Los participantes escucharon mensajes pregrabados con reseñas de productos a través de Alexa, Siri y Google Assistant, en versiones con valencia positiva o negativa. Las grabaciones se estandarizaron para asegurar consistencia en duración, estructura y estilo de entrega.

El análisis de datos combinó diversas técnicas estadísticas. Se utilizó el modelado de ecuaciones estructurales mediante mínimos cuadrados parciales (PLS-SEM) para evaluar las relaciones entre variables. Para explorar efectos de mediación y moderación, se aplicaron modelos PROCESS en SPSS y R, lo que permitió examinar efectos indirectos y términos de interacción. R también se utilizó para análisis complementarios, asegurando profundidad y coherencia analítica.

Los resultados muestran que el tipo de voz es un factor determinante en la formación de percepciones emocionales y sociales. Las voces que suenan más humanas aumentan la sensación de seguridad y la confianza, lo que se traduce en una mayor disposición a comprar. La influencia del tipo de voz sobre la intención de compra fue mayoritariamente indirecta, mediada por factores psicológicos como la respuesta emocional, la seguridad percibida, la atracción social y la confianza. Además, la voz ejerció un papel moderador, influyendo en la fuerza de relaciones clave, como el impacto de las actitudes previas o del boca a boca en las decisiones finales. Entre los tres asistentes analizados, la voz de Siri fue la que generó de forma más constante los resultados más favorables en estas variables.

Esta tesis revela que las compras por voz son mucho más que un intercambio funcional: constituyen una experiencia emocional y relacional. Los hallazgos ofrecen claves para desarrollar tecnologías de voz más humanas y emocionalmente inteligentes. Al reconocer la profundidad psicológica de hablar con una máquina, esta tesis nos invita a replantear cómo diseñamos las voces del futuro, no solo para informar, sino para conectar.

INTRODUCCIÓN

En los últimos años, la inteligencia artificial (IA) ha transformado la interacción humano-tecnología, pasando de interfaces basadas en texto a sistemas que procesan lenguaje natural, facilitando experiencias más intuitivas. Los asistentes virtuales han adquirido un rol central en tareas cotidianas, desde la atención al cliente hasta el comercio electrónico. Sin embargo, su falta de conexión humana sigue siendo un reto para su adopción, afectando la confianza del usuario, especialmente en compras online (Fakhimi et al., 2023). Para abordar este problema, emerge con fuerza el interés por la antropomorfización de la IA, es decir, el diseño de tecnologías que imiten rasgos humanos para mejorar su aceptación. Diversas investigaciones muestran que los usuarios tienden a antropomorfizar los chatbots, lo que fortalece la confianza y la facilidad de uso (Chinmulgund et al., 2023). No obstante, algunos estudios advierten que la antropomorfización excesiva puede generar expectativas poco realistas y plantear problemas en la percepción de las capacidades reales de estos sistemas (Deshpande et al., 2023).

Empresas y academia han explorado diversas estrategias para humanizar la IA. Algunas investigaciones han centrado su atención en el diseño visual de los chatbots, incorporando avatares con expresiones faciales que refuercen la sensación de cercanía y conexión emocional (Zhang & Yang, 2022), otros han analizado cómo el contenido del mensaje y el uso de lenguaje natural influyen en la percepción del usuario, permitiendo que la interacción con estos sistemas resulte más fluida y natural (Carter et al., 2024) y, aunque la antropomorfización puede lograrse a través de distintos elementos, la voz es uno de los más inmediatos y efectivos para una interacción natural con los usuarios (Fakhimi et al., 2023), ya que es un factor clave en la construcción de una experiencia más humana, dado su papel central en la comunicación y su capacidad para transmitir emociones.

La comunicación verbal no solo transmite información, sino también emociones, confianza y cercanía social (Fussell, 2002). Por este motivo, dotar a los asistentes virtuales de una voz natural y expresiva se ha convertido en una prioridad en la investigación y el desarrollo de tecnologías conversacionales. Ha quedado demostrado cómo la calidad de la voz influye en la presencia social percibida, es decir, en cómo de humano se percibe un asistente (Fakhimi et al., 2023) así, una voz cálida y expresiva genera mayor confianza, mientras que una robótica o monótona provoca distanciamiento (Chen & Park, 2021). En el ámbito del comercio electrónico, humanizar asistentes virtuales para potenciar la adopción de

las compras por voz es clave para mejorar la experiencia del cliente (Mari et al., 2024). Conscientes de ello, empresas como Amazon, Google y Apple han desarrollado tecnologías de voz avanzadas para hacer que sus asistentes sean más naturales y confiables. Sin embargo, a pesar de estos avances, la adopción de las compras por voz sigue siendo limitada en comparación con otros canales digitales, lo que indica que aún persisten barreras en su aceptación (Deshpande et al., 2023).

Uno de los principales motores detrás de la antropomorfización en asistentes virtuales es la necesidad de mejorar la experiencia del usuario y optimizar su interacción con la tecnología. Cuando un asistente de voz se percibe como más humano, la comunicación se vuelve más fluida, lo que reduce fricciones en la interacción, aumenta la satisfacción del usuario y refuerza su confianza en la tecnología (Chen & Park, 2021). En este contexto, la presencia social juega un papel clave, ya que permite que el usuario perciba al asistente no como un sistema automatizado, sino como un agente con intenciones y emociones, lo que favorece la generación de mayor confianza y engagement (Fakhimi et al., 2023). La literatura ha demostrado que los asistentes virtuales con una presencia social más fuerte no solo mejoran la experiencia de usuario, sino que también fomentan la aceptación de recomendaciones de productos y la realización de compras (Bawack et al., 2024; Davlembayeva et al., 2025). Además, la antropomorfización también incide en la percepción de seguridad, un aspecto crucial en el comercio electrónico. Un asistente de voz que suene más humano puede ayudar a reducir barreras psicológicas relacionadas con la privacidad y la seguridad en las transacciones online (Deshpande et al., 2023). No obstante, la relación entre la humanización de la voz, la seguridad percibida y la intención de compra aún requiere mayor estudio. Esto justifica la necesidad de investigar cómo el diseño vocal de los asistentes virtuales puede optimizar la confianza del usuario y potenciar la adopción del comercio por voz.

Si bien la antropomorfización de la IA puede lograrse a través de múltiples estrategias, esta tesis doctoral se centra en la voz como el principal factor de humanización en asistentes virtuales. A diferencia de otros elementos como la imagen o el lenguaje del sistema, la voz es el medio de comunicación más inmediato y natural para los humanos (Owren, 2011). Su impacto es aún más relevante en el comercio por voz, donde la interacción con el asistente no cuenta con una interfaz visual, convirtiendo la voz en el único canal de comunicación entre el usuario y el sistema (Fakhimi et al., 2023). Mientras que en entornos

gráficos los consumidores pueden apoyarse en la lectura de textos o en imágenes de productos, en los asistentes de voz la experiencia de compra depende exclusivamente de la información verbal proporcionada por el sistema. En este contexto, las características de la voz se convierten en un factor determinante para la confianza del usuario y su intención de compra, influyendo en la manera en que percibe la fiabilidad y credibilidad del asistente (Chen & Park, 2021).

Estudios previos han explorado la valencia del mensaje, es decir, si el contenido transmitido es positivo o negativo, y su impacto en la percepción del usuario (Chen & Xia, 2024). También se ha investigado cómo la inclusión de imágenes influye en la confianza y la percepción de humanidad en chatbots y asistentes virtuales con interfaces gráficas (Zhang & Yang, 2022). No obstante, en el caso de los asistentes de voz, el aspecto más determinante para mejorar la experiencia del usuario no es únicamente el contenido del mensaje, sino la forma en que se comunica. El tipo de voz del asistente puede marcar la diferencia entre una experiencia de compra percibida como confiable y fluida, o como fría y distante (Carter et al., 2024).

Por estas razones, esta tesis doctoral se centra en analizar cómo la voz impacta la percepción de humanidad, seguridad, confianza y la intención de compra en asistentes virtuales. A través de un enfoque empírico, se busca comprender qué características vocales generan mayor conexión emocional con el usuario y de qué manera estas influyen en la aceptación de la compra por voz en entornos digitales. Al profundizar en la relación entre la humanización de la voz y la intención de compra, esta investigación aporta conocimientos clave para el diseño de asistentes de voz más efectivos y alineados con las necesidades del consumidor digital.

El concepto de humanización en asistentes virtuales ha sido ampliamente debatido en la literatura sobre interacción humano-ordenador (HCI), donde se define como el proceso de dotar a la IA de características humanas para mejorar la experiencia del usuario y lograr una interacción más natural y efectiva (Simas & Ulbricht, 2024). Esta humanización se logra a través de distintos elementos, como la capacidad del asistente para procesar el lenguaje natural, la incorporación de avatares visuales y, en el caso de asistentes de voz, la calidad y expresividad de la voz utilizada. Mientras que en interfaces gráficas los usuarios pueden percibir humanidad a través de expresiones faciales o gestos de un avatar, en los asistentes de

voz el único canal de interacción es el lenguaje hablado. Esto hace que la presencia social percibida sea un factor determinante en la confianza y aceptación del usuario, ya que influye directamente en la forma en que la tecnología es interpretada y adoptada (Chen & Park, 2021).

La presencia social se define como el grado en que un sistema interactivo es percibido como un agente social con intenciones y emociones, en lugar de una simple herramienta mecánica (Skalski & Tamborini, 2007). Basado en el paradigma CASA (*Computers Are Social Actors*), este concepto sostiene que los humanos aplican normas sociales a la interacción con la tecnología, respondiendo a las máquinas como si fueran agentes sociales (Nass & Moon, 2000). Investigaciones previas muestran que los usuarios atribuyen características humanas a tecnologías con comportamientos sociales, como lenguaje coloquial, entonación expresiva o respuestas flexibles (Jensen et al., 2021). En el caso de los asistentes de voz, el diseño de la voz influye en la percepción de interacción con un agente "vivo" en lugar de un sistema automatizado. Este factor resulta fundamental en contextos de comercio electrónico, donde la confianza y la seguridad son aspectos esenciales. Para que los consumidores se sientan cómodos compartiendo información personal y tomando decisiones de compra basadas en la interacción con inteligencia artificial, es necesario que el asistente transmita fiabilidad y cercanía a través de su voz.

La antropomorfización, entendida como la atribución de características humanas a entidades no humanas, es un concepto clave en este estudio. En inteligencia artificial, se ha analizado en chatbots, robots sociales y asistentes virtuales para comprender cómo influye en la percepción y confianza del usuario (Deshpande et al., 2023). Existen distintos niveles de antropomorfización: pequeños detalles, como un nombre propio o una entonación cálida, pueden fortalecer la conexión emocional. Sin embargo, si es excesiva o artificial, puede generar rechazo, como explica la hipótesis del *Uncanny Valley* (Mori, 1970). Esta teoría sugiere que la aceptación de una entidad artificial aumenta con su similitud humana hasta cierto punto, pero si la semejanza es demasiado alta sin parecer completamente natural, puede causar incomodidad o rechazo. En el caso de los asistentes de voz, la voz es un elemento determinante en esta percepción: una voz demasiado robótica puede parecer fría y distante, mientras que una excesivamente humanizada puede generar sensación de falsedad o exageración, provocando una reacción negativa.

Si bien la humanización de la IA puede abordarse desde distintas dimensiones, esta

tesis se centra en la voz como su principal factor en asistentes virtuales. A diferencia de los elementos visuales o textuales, la voz transmite emociones de manera más inmediata y espontánea, lo que contribuye a generar una experiencia más inmersiva (Kraus, 2017). Investigaciones previas muestran que una voz con mayor carga emocional y matices expresivos aumenta la confianza del usuario y su disposición a seguir recomendaciones (Saffarizadeh et al., 2024). Sin embargo, aún hay un vacío en la literatura sobre su impacto en la intención de compra en comercio electrónico, destacando la necesidad de un análisis más profundo.

Además de la antropomorfización y la presencia social, la percepción de seguridad en asistentes de voz es un concepto clave en esta tesis doctoral. En el contexto del comercio electrónico, la seguridad percibida influye en la toma de decisiones del consumidor, especialmente al compartir información personal o realizar transacciones. Estudios previos sugieren que voces más cálidas y naturales generan mayor seguridad, mientras que voces artificiales pueden reducir la confianza en la tecnología (Sun & Liang, 2023). Sin embargo, la relación entre la voz del asistente y la percepción de seguridad ha sido poco explorada en la literatura, lo que convierte a esta investigación en una contribución relevante para comprender mejor este fenómeno.

Finalmente, es importante destacar que la humanización de los asistentes virtuales no solo se limita a la experiencia de usuario, sino que también tiene implicaciones directas en la adopción y fidelización de estas tecnologías en el mercado. Empresas como Google, Amazon y Apple han desarrollado asistentes de voz con diferentes estrategias de diseño vocal, intentando equilibrar la funcionalidad con una sensación de cercanía y confianza para el usuario. Sin embargo, la manera en que los consumidores perciben estas diferencias y cómo influyen en su comportamiento de compra sigue siendo un área de estudio en evolución. A través de esta tesis, se busca aportar evidencia empírica sobre cómo la voz de un asistente virtual puede impactar no solo en la interacción inmediata con el usuario, sino también en su confianza, seguridad percibida e intención de compra a largo plazo.

La adopción de asistentes de voz en el comercio electrónico ha experimentado un crecimiento acelerado en los últimos años, impulsada por los avances en inteligencia artificial y la creciente disponibilidad de dispositivos compatibles. Sin embargo, su uso en compras sigue siendo limitado frente a otros canales digitales. Aunque los consumidores recurren a

estos sistemas para buscar información o comparar precios, son pocos los que finalizan el proceso de compra a través de ellos (Huh et al., 2023). Esta brecha sugiere que aún existen barreras en la aceptación de la compra por voz, siendo la falta de confianza en la interacción con estos sistemas un factor clave en la disposición del usuario a realizar transacciones mediante este canal. En este contexto, la humanización del asistente de voz es clave para generar confianza. Estudios muestran que los consumidores tienden a confiar más en asistentes con voces cálidas y naturales, mientras que aquellos con voces robóticas o mecánicas suelen percibirse como impersonales y menos fiables (Zaharia & Würfel, 2021). Además, la percepción de presencia social, entendida como la sensación de que el asistente tiene intenciones y emociones, influye en la forma en que los usuarios interactúan con la tecnología, afectando su nivel de confianza y su predisposición a utilizar estos sistemas (Rhee & Choi, 2020). A pesar de estos hallazgos, la relación entre el tipo de voz, la percepción de seguridad y la intención de compra sigue siendo poco estudiada. Esta brecha en la literatura subraya la necesidad de una investigación más profunda que permita comprender de manera más precisa cómo la humanización de la voz en asistentes virtuales puede impactar la adopción del comercio por voz.

Además de la confianza, la percepción de seguridad es un desafío en la compra por voz. A diferencia de las plataformas visuales, donde el consumidor accede a información detallada antes de decidir, la compra por voz depende solo de la información verbal del asistente, lo que puede generar incertidumbre, ya que la ausencia de referencias visuales dificulta la confianza en la tecnología (Bawack et al., 2021). Aunque se han realizado estudios en esta área, aún no se ha explorado en profundidad cómo las características vocales influyen en la confianza del usuario y su disposición a completar una compra. A estos factores se le suma la falta de conexión emocional con los asistentes virtuales. A diferencia de otras formas de comercio electrónico, los asistentes de voz pueden parecer impersonales, reduciendo la sensación de interacción humana y afectando su uso continuo. Si bien se ha sugerido que la capacidad de un asistente para transmitir emociones podría mejorar la confianza y la satisfacción del usuario (Huh et al., 2023), aún no se ha investigado en profundidad qué emociones serían más efectivas ni cómo su presencia en la voz del asistente impacta la intención de compra. Esta brecha en la literatura refuerza la necesidad de comprender mejor el papel de la expresividad vocal en la percepción del usuario y en la adopción del comercio por voz.

Desde una perspectiva aplicada, esta investigación responde a la creciente adopción del comercio por voz, cuyo mercado vinculado a altavoces inteligentes alcanzará aproximadamente 29,65 mil millones de dólares en ingresos globales para el año 2029 (Statista Consumer Market Insights, 2024). Sin embargo, la reticencia de los consumidores a confiar plenamente en estos sistemas sigue siendo una barrera para su adopción masiva. Grandes compañías como Amazon, Google y Apple han invertido significativamente en la mejora de sus asistentes de voz con el objetivo de convertirlos en plataformas de compra confiables y eficientes. No obstante, la falta de estudios detallados sobre cómo optimizar la experiencia del usuario a través del diseño vocal dificulta la implementación de estrategias efectivas para aumentar la conversión en estos canales.

Además, esta tesis se alinea con las prioridades del Marketing Science Institute (MSI) 2022-2024, en particular con la prioridad 2.4: Atención, Engagement y Experiencia del Cliente. Esta línea de investigación busca comprender cómo los consumidores procesan información, interactúan con tecnologías emergentes y responden a estímulos digitales. El estudio de la percepción de humanidad en la voz de los asistentes virtuales aporta nuevos conocimientos sobre su impacto en la confianza del consumidor y la intención de compra. El MSI también destaca la necesidad de mejorar la medición de mediadores psicológicos, como la confianza y la seguridad percibida, en la interacción con tecnologías y sus efectos comerciales (Cowan et al., 2021). En este sentido, esta tesis evalúa cómo la percepción de humanidad en la voz influye en la confianza del usuario y su disposición a seguir recomendaciones automatizadas. Asimismo, el MSI enfatiza la optimización de la experiencia del cliente en entornos digitales mediante tecnologías que no solo mejoren la eficiencia de compra, sino que también fortalezcan la conexión emocional del usuario. En esta línea, esta tesis analiza el papel de la expresividad vocal y la personalización en la generación de confianza y seguridad en el comercio electrónico.

Por último, la integración de asistentes virtuales en el comercio electrónico plantea preguntas fundamentales sobre cómo los consumidores perciben y valoran las interacciones automatizadas. El comercio por voz ha transformado la forma en que los consumidores interactúan con las plataformas digitales, permitiéndoles realizar compras y recibir recomendaciones a través de asistentes virtuales. No obstante, su adopción aún enfrenta barreras significativas, que no solo dependen de la funcionalidad tecnológica, sino también de factores psicológicos, emocionales y cognitivos que influyen en la experiencia del usuario. La

efectividad de los asistentes de voz dependerá de su capacidad para generar interacciones empáticas, atractivas, confiables y, en definitiva, más humanas.

Por ello, comprender qué elementos determinan la confianza, la seguridad percibida y la disposición del consumidor a aceptar recomendaciones automatizadas se ha convertido en una prioridad tanto para la investigación académica como para el desarrollo de estrategias comerciales eficaces en entornos digitales (Luo et al., 2023; McLean & Osei-Frimpong, 2019). El desarrollo de esta tesis se alinea con la agenda del MSI al abordar estas cuestiones desde un enfoque basado en la psicología del consumidor y la interacción humano-ordenador, contribuyendo al diseño de estrategias que mejoren la experiencia del usuario y optimicen la adopción del comercio por voz.

En conclusión, el problema central de esta tesis doctoral radica en la necesidad de comprender cómo la voz influye en la experiencia de compra por voz, particularmente en lo que respecta a la confianza, la seguridad y la conexión emocional con el usuario. Si bien existen estudios previos sobre la importancia de la humanización en asistentes virtuales (Chérif & Lemoine, 2019), aún no se ha explorado a fondo cómo distintos tipos de voz pueden optimizar la percepción de humanidad y mejorar la intención de compra en contextos de comercio electrónico. Esta tesis busca aportar nuevos conocimientos en este ámbito, proporcionando información clave para el diseño de asistentes de voz más efectivos y alineados con las expectativas del consumidor digital.

Teniendo en cuenta las brechas en la literatura sobre interacción humano-ordenador y comercio online, así como las prioridades de investigación formuladas por el Marketing Science Institute 2022-2024, esta tesis se organiza en cuatro estudios principales. Cada uno aborda un objetivo de investigación clave, centrado en responder cuestiones aún no resueltas sobre la humanización de los asistentes de voz y su impacto en la confianza, la seguridad y la experiencia de compra en entornos digitales.

El objetivo principal de esta tesis es analizar la aceptación de los asistentes virtuales de voz en la compra online y evaluar cómo las características de la voz influyen en la seguridad percibida, la transmisión de emociones, la confianza y la intención de compra. A partir de estos análisis, se busca identificar qué atributos vocales resultan más eficaces para optimizar la interacción con los usuarios. Para alcanzar este propósito, la investigación se estructura en cuatro capítulos, cada uno de los cuales responde a un objetivo de investigación

específico.

- **Objetivo de investigación 1**: *Analizar los factores que influyen en la aceptación del comercio por voz, integrando modelos teóricos como el Technology Acceptance Model (TAM) y la Uses and Gratifications Theory (UGT) para comprender la intención de adopción de los asistentes de voz en el contexto de las compras online.*

El crecimiento del comercio por voz ha generado un interés creciente en entender qué factores influyen en la aceptación de esta tecnología por parte de los consumidores. Si bien el comercio electrónico ha evolucionado con la incorporación de asistentes de voz, la tasa de adopción sigue siendo baja en comparación con otros medios digitales. Factores como la percepción de seguridad, la utilidad percibida, el disfrute del usuario y la facilidad de uso han sido identificados en la literatura como elementos clave en la adopción de nuevas tecnologías. Sin embargo, su impacto en la intención de compra mediante asistentes de voz sigue sin estar completamente definido. Este objetivo busca proporcionar un marco teórico integral para analizar la adopción del comercio por voz, combinando el TAM y la UGT para evaluar cómo los consumidores perciben la utilidad, la seguridad y la satisfacción en el uso de asistentes de voz en compras online.

El capítulo 1 de esta tesis aborda este objetivo mediante un análisis detallado de los factores que determinan la aceptación del comercio por voz. Utilizando un enfoque basado en el TAM y la UGT, se analizan las variables clave que influyen en la intención de adopción de los consumidores.

Este capítulo es una extensión del trabajo de: Calahorra-Candao, G., & Martín-de Hoyos, M. J. (2024). From Typing to Talking: Unveiling AI's Role in the Evolution of Voice Assistant Integration in Online Shopping. Information, 15(4), 202.

Si bien el primer capítulo analiza los factores generales que influyen en la aceptación del comercio por voz, no profundiza en el impacto de la voz en la experiencia del usuario. Por ello, el siguiente paso es examinar cómo la humanización de los asistentes virtuales a través de la voz afecta la percepción de seguridad y la intención de compra. Este aspecto resulta fundamental en la interacción usuario-tecnología y puede ser determinante para el éxito del comercio por voz.

- **Objetivo de investigación 2**: *Examinar el efecto de la antropomorfización en la percepción de seguridad de los asistentes de voz y cómo esta percepción influye en la intención de compra mediante estos dispositivos.*

Uno de los principales desafíos en la adopción de la compra por voz es la percepción de seguridad. Los consumidores necesitan confiar en que los asistentes virtuales protegerán su información y ofrecerán una experiencia de compra confiable. La literatura ha demostrado que los asistentes con características más antropomórficas, como una voz más humana y expresiva, pueden generar una mayor sensación de presencia social y confianza. Sin embargo, el impacto específico de la humanización vocal en la seguridad percibida y su efecto en la intención de compra sigue sin estar completamente definido. Este estudio analiza cómo la humanización de la voz en los asistentes virtuales influye en la percepción de seguridad del usuario y de qué manera esta percepción impacta en su intención de realizar compras mediante comandos de voz.

Este capítulo es una extensión del trabajo de: Calahorra-Candao, G., & Martín-de Hoyos, M. J. (2024). The effect of anthropomorphism of virtual voice assistants on perceived safety as an antecedent to voice shopping. Computers in Human Behavior, 153, 108124.

El capítulo 2 de esta tesis se centra en el impacto de la antropomorfización vocal en la percepción de seguridad y la intención de compra en asistentes de voz. Este capítulo examina cómo la humanización de los asistentes mediante la voz influye en la percepción de seguridad, lo que a su vez puede favorecer la aceptación de la compra por voz. La confianza en la tecnología no solo depende de la seguridad percibida, sino también de la capacidad de los asistentes para generar una conexión emocional con los usuarios. En este sentido, la percepción de humanidad y la presencia social pueden desempeñar un papel fundamental en la adopción y el uso continuo de estas tecnologías.

- **Objetivo de investigación 3:** *Analizar el impacto del diseño emocional en la confianza del usuario en asistentes de voz, evaluando la capacidad de las distintas voces para transmitir emociones básicas y cómo influyen en la percepción de humanidad.*

En el capítulo 3 se analizan específicamente las diferencias entre tipos de voz para determinar qué características vocales generan mayor confianza y conexión con el usuario.

La creciente integración de asistentes de voz en la vida cotidiana ha llevado a un interés creciente en comprender cómo la emoción percibida en sus voces afecta la interacción humano-ordenador. La literatura ha demostrado que los asistentes de voz más expresivos pueden generar una mayor sensación de presencia social y conexión emocional con los usuarios. Sin embargo, aún no está claro qué tipo de emociones son más efectivas para generar confianza y compromiso en entornos comerciales.

Este objetivo de investigación busca profundizar en el impacto del diseño emocional de los asistentes de voz en la percepción del usuario. En este contexto, el capítulo 3 amplía el análisis al explorar el papel de las emociones en la interacción con asistentes de voz, estudiando las diferencias existentes entre tres tipos de voces: Alexa, Google y Siri. En particular, se investiga si las características de cada voz son capaces de transmitir emociones de manera diferenciada y cómo esto influye en la percepción de estos sistemas como agentes sociales con características humanas. Además, se analiza el posible efecto de la valencia del mensaje, es decir, si expresa emociones tanto positivas como negativas.

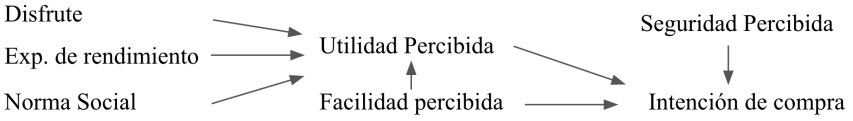
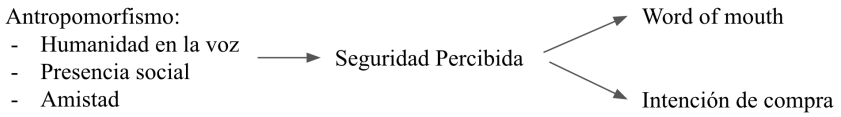
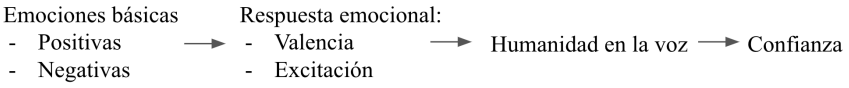
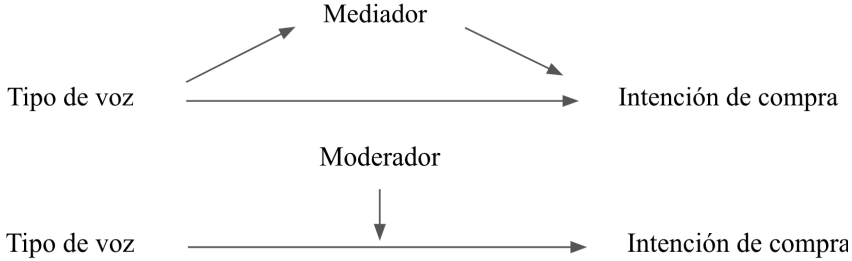
Si bien el capítulo 3 explora cómo la expresividad emocional de cada voz de los asistentes transmite emociones que afectan a la confianza del usuario y su percepción de humanidad, aún queda por analizar cómo estas percepciones influyen en el comportamiento del consumidor en entornos de comercio por voz. El capítulo 4 amplía este análisis, evaluando cómo el tipo de voz de los asistentes virtuales influye en la intención de compra. En este estudio se comparan las tres voces seleccionadas para analizar su impacto en la disposición del usuario a realizar compras mediante asistentes de voz.

- **Objetivo de investigación 4**: *Evaluar cómo la tipología de voz de los asistentes virtuales influye en la percepción de confianza, seguridad, en la intención de recomendar y la intención de compra.*

Dado que la confianza y la seguridad percibida son factores clave en la aceptación de la compra por voz, es fundamental comprender cómo distintos tipos de voz afectan de forma directa e indirecta a la percepción del usuario y su predisposición a realizar transacciones mediante asistentes virtuales de forma diferente. Mientras que voces más naturales pueden generar mayor confianza y credibilidad, voces más sintéticas o robóticas podrían reducir la aceptación del usuario. Este estudio busca determinar qué características vocales optimizan la experiencia de compra y facilitan la conversión en plataformas de comercio electrónico.

La Tabla 0.1 resume el objetivo principal y los específicos de esta tesis doctoral, presentando los modelos conceptuales de cada estudio y estableciendo su relación con las prioridades de investigación no resueltas formuladas por el Marketing Science Institute 2022-2024.

Tabla 0.1. Modelos conceptuales

Prioridades de investigación del MSI 2022-2024	**Objetivo principal**: Analizar cómo las características vocales de los asistentes virtuales impactan la experiencia del consumidor en el comercio digital.	Modelos conceptuales
> ¿Cuál es la mejor manera de medir la experiencia del cliente? > ¿Cómo influyen distintos factores en la adopción de nuevas tecnologías en el comportamiento del consumidor?	**Capítulo 1**: Analizar los factores clave que influyen en la adopción del comercio por voz, integrando modelos teóricos como el TAM y la UGT para comprender los impulsores de la adopción de asistentes de voz en las compras online.	
> ¿Cómo puede el uso de la tecnología mejorar o perjudicar la experiencia del cliente, equilibrando el deseo de conexión humana auténtica con la necesidad de un servicio inmediato y sin fricciones?	**Capítulo 2**: Examinar cómo la antropomorfización de los asistentes de voz afecta la percepción de seguridad y la confianza del usuario, y cómo estos factores influyen en la disposición de los consumidores a realizar compras mediante asistentes de voz.	
> ¿Cómo responden los consumidores a diferentes tipos de contenido y formatos en las interacciones online?	**Capítulo 3**: Analizar el papel de la voz en la transmisión de emociones que humanicen la interacción, estudiando cómo diferentes estilos de voz y entonaciones influyen en la generación de confianza.	
> ¿Cómo pueden optimizarse las técnicas de comunicación persuasiva para aumentar la respuesta del consumidor a los mensajes de marketing? > ¿Cómo podemos alinear mejor las estrategias de marketing y comunicación con las preferencias del consumidor?	**Capítulo 4**: Evaluar los efectos directos e indirectos del tipo de voz en la intención de compra así como los efectos mediadores de la confianza, seguridad, humanidad y respuesta emocional.	

Referencias

Bawack, R. E., Bonhoure, E., & Mallek, S. (2024). Why would consumers risk taking purchase recommendations from voice assistants? *Information Technology & People*.

Bawack, R. E., Wamba, S. F., & Carillo, K. D. A. (2021). Adoption of Smart Speakers for Voice Shopping. En F. Ceci, A. Prencipe, & P. Spagnoletti (Eds.), *Exploring Innovation in a Digital World* (pp. 21-35). Springer International Publishing. https://doi.org/10.1007/978-3-030-87842-9_3

Carter, O. B., Loft, S., & Visser, T. A. (2024). Meaningful communication but not superficial anthropomorphism facilitates human-automation trust calibration: The Human-Automation Trust Expectation Model (HATEM). *Human Factors*, *66*(11), 2485-2502.

Chen, J., & Xia, S. (2024). Are online users influenced by what other users say? Meta-analyzing the cognitive, emotional, and behavioral impact of online comment valence. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *18*(5).

Chen, Q., & Park, H. (2021). How anthropomorphism affects trust in intelligent personal assistants. *Ind. Manag. Data Syst.*, *121*, 2722-2737. https://doi.org/10.1108/imds-12-2020-0761

Chérif, E., & Lemoine, J.-F. (2019). Anthropomorphic virtual assistants and the reactions of Internet users: An experiment on the assistant's voice. *Recherche et Applications En Marketing (English Edition)*, *34*(1), 28-47. https://doi.org/10.1177/2051570719829432

Chinmulgund, A., Khatwani, R., Tapas, P., Shah, P., & Sekhar, R. (2023). Anthropomorphism of AI based chatbots by users during communication. *2023 3rd International Conference on Intelligent Technologies (CONIT)*, 1-6.

Cowan, K., Javornik, A., & Jiang, P. (2021). Privacy concerns when using augmented reality face filters? Explaining why and when use avoidance occurs. *Psychology &*

Marketing, *38*(10), 1799-1813. https://doi.org/10.1002/mar.21576

Davlembayeva, D., Chari, S., & Papagiannidis, S. (2025). Virtual influencers in consumer behaviour: A social influence theory perspective. *British Journal of Management*, *36*(1), 202-222.

Deshpande, A., Rajpurohit, T., Narasimhan, K., & Kalyan, A. (2023). Anthropomorphization of AI: opportunities and risks. *arXiv preprint arXiv:2305.14784*.

Fakhimi, A., Garry, T., & Biggemann, S. (2023). The Effects of Anthropomorphised Virtual Conversational Assistants on Consumer Engagement and Trust During Service Encounters. *Australasian Marketing Journal*, *31*, 314-324. https://doi.org/10.1177/14413582231181140

Fussell, S. R. (2002). The verbal communication of emotion: Introduction and overview. En *The verbal communication of emotions* (pp. 1-16). Psychology Press.

Huh, J., Whang, C., & Kim, H.-Y. (2023). Building trust with voice assistants for apparel shopping: The effects of social role and user autonomy. *Journal of Global Fashion Marketing*, *14*(1), 5-19.

Jensen, T., Khan, M. M. H., Fahim, M. A. A., & Albayram, Y. (2021). Trust and anthropomorphism in tandem: The interrelated nature of automated agent appearance and reliability in trustworthiness perceptions. *Proceedings of the 2021 ACM Designing Interactive Systems Conference*, 1470-1480.

Kraus, M. W. (2017). Voice-only communication enhances empathic accuracy. *American Psychologist*, *72*(7), 644.

Luo, B., Lau, R. Y., & Li, C. (2023). Emotion-regulatory chatbots for enhancing consumer servicing: An interpersonal emotion management approach. *Information & Management*, *60*(5), 103794.

Mari, A., Mandelli, A., & Algesheimer, R. (2024). Empathic voice assistants: Enhancing consumer responses in voice commerce. *Journal of Business Research*, *175*, 114566.

McLean, G., & Osei-Frimpong, K. (2019). Hey Alexa … examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in Human Behavior*, *99*, 28-37. https://doi.org/10.1016/j.chb.2019.05.009

Mori, M. (1970). The uncanny valley. *Energy*, *7*(4), 33-35.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of social issues*, *56*(1), 81-103.

Owren, M. J. (2011). Human voice in evolutionary perspective. *Acoust. Today*, *7*(24), 10-1121.

Rhee, C. E., & Choi, J. (2020). Effects of personalization and social role in voice shopping: An experimental study on product recommendation by a conversational voice agent. *Computers in Human Behavior*, *109*, 106359. https://doi.org/10.1016/j.chb.2020.106359

Saffarizadeh, K., Keil, M., Boodraj, M., & Alashoor, T. (2024). "My Name is Alexa. What's Your Name?" The Impact of Reciprocal Self-Disclosure on Post-Interaction Trust in Conversational Agents. *Journal of the Association for Information Systems*, *25*(3), 528-568.

Simas, G., & Ulbricht, V. R. (2024). Human-AI interaction: An analysis of Anthropomorphization and user engagement in conversational agents with a focus on ChatGPT. *Intelligent Human Systems Integration (IHSI 2024): Integrating People and Intelligent Systems*, *119*(119).

Skalski, P., & Tamborini, R. (2007). The role of social presence in interactive agent-based persuasion. *Media psychology*, *10*(3), 385-413.

Statista Consumer Market Insights. (2024, junio 28). *Smart speaker market revenue from 2019 to 2029 (in billion U.S. dollars)*. https://www.statista.com/forecasts/1367973/smart-speaker-market-revenue-worldwide

Sun, Y., & Liang, Q. (2023). Research on the Relationship between Self-AI Connection and Customer engagement–Based on the Analysis of Customer AI Trust and AI

Anthropomorphism. *Proceedings of the 2023 6th International Conference on Information Management and Management Science*, 48-53.

Zaharia, S., & Würfel, M. (2021). Voice Commerce-Studying the Acceptance of Smart Speakers. *Proceedings of the 3rd International Conference on Human Interaction and Emerging Technologies: Future Applications (IHIET 2020)*, 449-454.

Zhang, A., & Yang, Q. (2022). To be human-like or machine-like? An empirical research on user trust in AI applications in service industry. *2022 8th International Conference on Automation, Robotics and Applications (ICARA)*, 9-15.

STUDY 1

Acceptance and Use of Voice Assistants

Study 1 is an extension of the work: Calahorra-Candao, G., & Martín-de Hoyos, M. J. (2024). From Typing to Talking: Unveiling AI's Role in the Evolution of Voice Assistant Integration in Online Shopping. Information, 15(4), 202.

1.1. Introduction

The Echo Dot, a pioneering voice assistant (VA) powered by Alexa, has consistently led Amazon's Black Friday sales, showcasing the escalating popularity of VAs in diverse spheres (Amazon Staff, 2022). These AI-powered devices, adept at simulating human speech, have evolved beyond rudimentary tasks like scheduling and music playback. They are increasingly integrated into household appliances and commercial settings, underscoring a significant shift in consumer interaction with technology. Prominent brands such as Domino's Pizza and Dunkin' Donuts have adopted VAs for customer orders, illustrating a growing trend in commercial utilization (Dousay & Hall, 2018; Huang & Benyoucef, 2013).

The VA market is experiencing substantial global growth, projected to reach $27.16 billion by 2026, a significant increase from $10.7 billion in 2020 (Mordor Intelligence, 2021). This is paralleled by the estimated shipment of over 409.4 million smart speakers worldwide by 2025 (Loup Ventures, 2023). However, despite this rapid expansion, the application of VAs in making purchases is still nascent. Only 15% of U.S. smart-speaker users routinely use voice commands for shopping (Statista Staff, 2022). The COVID-19 pandemic catalyzed a 72% increase in smart speaker purchases in North America, suggesting an untapped potential in the voice shopping domain (Astound Commerce, 2023).

While the proliferation of VAs in both personal and commercial realms is well-documented, a striking gap persists in the literature, particularly their usage throughout the purchasing process. Studies to date have extensively explored VA acceptance for initial interactions, such as product inquiries or adding items to a cart, highlighting the importance of trust and perceived human-likeness in this context (Huh et al., 2023). However, they fall short in examining the role of VA acceptance in subsequent stages, including final purchasing decisions. This limitation indicates that while voice AI adoption increases browsing and spending, the actual completion of transactions through voice shopping appears to be less prevalent, suggesting a halt before the final transaction stage (Sun et al., 2019).

1.2. Voice shopping acceptance literature review

Voice shopping, a revolutionary method powered by voice-controlled devices such as smart speakers and smartphones, offers a seamless interface for consumers. These interfaces,

featured in devices with and without screens, like the Echo Dot by Amazon and the HomePod by Apple, signify a transformative shift in shopping experiences. This relatively new phenomenon has captured academic attention, with studies suggesting its potential for speed, seamlessness, and higher conversion rates in sales compared to traditional shopping methods (Simms, 2019).

The current landscape of voice shopping is characterized by a notable trend towards increased interest and acceptance, primarily fueled by the widespread availability of smart speakers and voice assistants. This growth trajectory is significantly influenced by the efficiency and convenience offered by these technologies (Huh et al., 2023; Zaharia & Würfel, 2021).

A critical aspect of voice shopping's evolution is the role of psychological factors that shape consumer attitudes. Studies indicate that social norm, perceived safety or performance expectancy are some of the pivotal factors in determining customer experience and their propensity to engage with voice shopping platforms (Bawack et al., 2021a). These factors not only influence the decision-making process in purchasing but also impact customer satisfaction and their willingness to adopt this technology for shopping (Celik, 2016).

When consumers search for product information, read reviews, or evaluate purchasing options, perceived safety often emerges as a major barrier to adoption. Concerns about data privacy and the potential risks associated with voice shopping platforms can significantly reduce users' willingness to engage with these technologies. Research has consistently shown that perceived risk is a strong inhibitor of technology adoption, especially when personal or sensitive information is involved. For instance, Liébana-Cabanillas et al., (2013) found that higher perceived risk directly decreases consumers' intention to adopt digital services, emphasizing the critical role of trust and perceived security in facilitating acceptance. These insights underscore the importance of proactively addressing privacy and security concerns to build user confidence. By enhancing trust, platforms can foster greater acceptance of voice-based shopping. Previous research further reinforces the centrality of these concerns, highlighting their strong influence on consumer attitudes toward the adoption of emerging shopping technologies (Celik, 2016; Zaharia & Würfel, 2021)

Another notable factor in the advancement of voice shopping is the impact of personalization and the social role of voice assistants. The way these assistants are perceived,

whether as a friend or a secretary, and how well they cater to individual preferences, can substantially affect consumer attitudes. Enhanced personalization and a relatable social role can significantly improve the shopping experience, leading to higher satisfaction and increased purchase intentions (Rhee & Choi, 2020).

While voice shopping is gaining momentum, it still has considerable potential for wider acceptance. Addressing existing challenges such as improving the technical sophistication of voice assistants, mitigating privacy risks, and enhancing user autonomy is essential. By focusing on these areas, there is an opportunity to boost consumer confidence, leading to increased acceptance and integration of voice shopping into everyday commerce (Chung et al., 2022; Huh et al., 2023).

In the exploration of voice shopping intention, a comprehensive theoretical approach is essential to fully grasp the underlying dynamics of consumer behavior in this emerging field. Several studies have adopted frameworks such as the Unified Theory of Acceptance and Use of Technology (UTAUT) and the TAM, focusing on critical factors like perceived ease of use, usefulness, social influence, and hedonic motivation. These elements are instrumental in shaping consumer perceptions and acceptance of voice shopping technology (Adolphs & Zaharia, 2021; Park, 2010).

However, it is noteworthy that there is no single, universally adopted theoretical approach to investigate voice shopping intention. The field is marked by a diversity of theories, with no consensus on which one offers the most comprehensive understanding. This range of theoretical approaches underscores the complexity and multifaceted nature of studying voice shopping intention. It highlights the need for a cohesive and integrated theoretical framework that can encapsulate the various dimensions of consumer behavior in voice shopping. Evidence of this can be found in Table 1.1, which presents a literature review in the respective studies, compiling some of the latest research in this area. This table showcases the theories utilized in each study, and from this review, we can ascertain that there is no single theory capable of fully explaining the intention behind voice shopping. This compilation underscores the diversity of theoretical approaches in the field, highlighting the complexity and multifaceted nature of studying consumer behavior in the context of voice shopping.

Table 1.1. Voice shopping acceptance studies.

Authors and year	Theory Used	Conclusions
Zaharia & Würfel (2021)	Unified Theory of Acceptance and Use of Technology 2 (UTAUT 2)	Factors influencing willingness to use smart speakers include hedonic motivation, performance expectancy, and negative effect of perceived risk.
Huh et al. (2023)	Media Equation Theory	Effects of social role and user autonomy on perceived human-likeness, trust, and purchase intention in voice shopping.
Bawack et al. (2021a)	Goldberg's Big Five Factors, Theory of Reasoned Action (TRA-privacy)	Impact of personality, trust, privacy concerns on customer experience performance in voice shopping.
Rhee & Choi (2020)	Dual Modes of Information-Processing Models	Persuasion mechanism in product recommendations by voice-based conversational agents.
Celik (2016)	Unified Theory of Acceptance and Use of Technology (UTAUT)	Influence of anxiety on the customer adoption of voice shopping within the UTAUT framework.
Gorgijevski et al. (2019)	Voice Behavior Theory, Attention-Based View of the Firm	Influence of initiative selling tactics on the acceptance of subsidiary initiatives by MNC headquarters.
Kun & Xing (2016)	Pluralist Industrial Relation View, Unary Human Resource Management Theory	Inspection and integration of employee voice research in view of its nature and effects.
Whang & Song (2019)	Epley et al.'s (2007) Anthropomorphism Model	Impact of perceived human likeness of voice assistants on purchase intention in voice shopping.
Natarajan et al. (2017)	Extended Technology Acceptance Model (TAM), Theory of Diffusion of Innovations (DOI)	Understanding the intention to use mobile shopping applications and its influence on price sensitivity.

Madlberger & O'Neill (2020)	Technology Acceptance Model, Unified Theory of Acceptance and Use of Technology, Social Exchange Theory, Privacy Concerns	Exploring antecedents of voice-activated assistant usage intention (Perceived Usefulness, Perceived Ease of Use, Social Influence, Trust and Security concerns).
Park et al. (2022)	CASA (Computers Are Social Actors)	Interaction with advertisements increases brand and product recognition compared to non-interactive advertisements. Contextual relevance enhances brand and product recognition, but talker variability does not affect it.
Huh & Kim (2022)	Uses and Gratifications Theory	The study focused on four gratification dimensions in voice shopping (life efficiency, entertainment, social presence, affordance) and found these dimensions, along with satisfaction, impact purchase intention.
Calahorra-Candao & Martín-de Hoyos (2024)	Communication privacy management (CPM) theory	The study concluded that the anthropomorphism of virtual voice assistants significantly influences perceived safety, which is a critical factor affecting consumer engagement in voice shopping.

Source: Own elaboration

Continuing with this analysis, it is observed that the application of the TAM and the UGT presents a combined framework particularly suited for examining voice shopping. While TAM provides insights into the functional attributes and perceived ease of use of voice shopping, UGT contributes a valuable perspective by focusing on the psychological engagement and experiential aspects that voice shopping platforms offer. This dual-theoretical approach enables a comprehensive analysis, encompassing both the practical utility and the experiential facets of voice shopping, thus offering a more complete understanding of consumer intentions in the context of this innovative shopping medium. The decision to employ TAM and UGT as the primary theoretical frameworks for studying voice shopping intention is supported by their specific applicability and comprehensiveness

compared to other theories in this domain.

The Technology Acceptance Model (TAM), initially formulated by Davis (1986) and further developed in 1989, is deeply entrenched in the principles of behavioral psychology, specifically drawing from the Theory of Reasoned Action (TRA) by Ajzen (1980). Originally designed to assess computer-based information systems, TAM has shown remarkable adaptability and has evolved to encompass a wide range of technological domains, including mobile commerce and VAs. This evolution highlights TAM's versatility and its continued relevance in modern research contexts, especially in the exploration of emerging technologies such as voice shopping (Venkatesh & Davis, 2000).

Particularly noteworthy in TAM's application to voice shopping is its focus on two central factors: perceived ease of use and perceived usefulness. These elements are critical in the context of voice shopping, where the simplicity of user interaction and the practical advantages offered — such as time efficiency and convenience — play a pivotal role in user acceptance (Zarmpou et al., 2012). Unlike broader models that might encompass a wider range of variables, TAM's focused constructs offer clear and direct predictors of technology acceptance. This specificity makes TAM an ideal framework for investigating the specific technological context of voice shopping, as it directly addresses the factors most influential in determining user acceptance and acceptance of this shopping method. The model's evolution and its application in the context of voice shopping demonstrate its flexibility and capacity to provide valuable insights into user behavior and technology acceptance in the rapidly evolving landscape of digital commerce.

Despite its widespread application in fields such as e-commerce, online banking, and e-learning environments, TAM has faced criticism for its limited scope in explaining the acceptance and usage of specific, innovative technologies such as voice shopping. Critics point out that TAM tends to overlook external contextual factors, including social influences, individual user differences, and situational variables, which can significantly impact technology acceptance behaviors (Bagozzi, 2007). This limitation necessitates additional theoretical support to provide a more nuanced understanding of user behavior, especially in the context of rapidly evolving technologies like voice assistants and voice shopping platforms (Venkatesh et al., 2003). Addressing this gap, our study integrates the Uses and Gratifications Theory (UGT) with TAM to gain a more comprehensive insight into voice

shopping acceptance.

UGT, or the Uses and Gratifications Theory, presents a valuable dimension in the study of media use by delving into the psychological and experiential aspects (Katz et al., 1973). UGT is particularly relevant for understanding the personal and emotional gratifications that users seek from modern media, including voice shopping platforms. This theory encompasses aspects such as entertainment, information, and personal identity, which are crucial in determining the appeal of these platforms (Rhee & Choi, 2020). The focus of UGT on user-centric motivations offers a deeper understanding of why consumers may prefer voice shopping to other methods. This perspective is crucial as it addresses aspects that might not be as effectively captured by theories focusing solely on the functional attributes of technology.

In the context of the digital era, where user motivations transcend mere functional aspects of technology, UGT's relevance is further accentuated. It explores elements inherent to the individual, like personal preferences, psychological needs, and social influences. These factors are pivotal in understanding why individuals opt for specific technologies over others. UGT's approach to understanding media use goes beyond the technological features to uncover the underlying motivations and needs of users. This makes it a particularly powerful tool in the study of voice shopping, where personal gratifications and individual experiences play a significant role in technology acceptance and use.

Thus, UGT complements more technically focused frameworks by providing a comprehensive view of user behavior in the context of digital technologies. Its emphasis on personal gratifications and psychological underpinnings provides a holistic understanding of consumer behavior, crucial for effectively analyzing and predicting the success of technologies like voice shopping in today's digital landscape.

For instance, research applying UGT in diverse contexts like internet use, social media, and chatbots demonstrates that factors like personal enjoyment, the need for social connection, and concerns about privacy are pivotal in driving technology usage. These factors are less about the technical capabilities of the platforms and more about the personal gratifications users seek. Studies on online platforms, social media or chatbots underscore this perspective (Cheng & Jiang, 2020; Naqvi et al., 2020). They highlight that while technological attributes are important, it's the individual's subjective experience, such as the

pleasure derived from using the technology, the desire for social interaction, or concerns about personal data security, that significantly influences their intention to use the technology. This shift in focus towards user-centric motivations aligns with the broader trend in technology research, emphasizing the importance of understanding users' internal drivers and social contexts to fully grasp their interaction with digital technologies.

In the context of voice shopping, combining the Technology Acceptance Model (TAM) with the Uses and Gratifications Theory (UGT) offers a robust framework for understanding consumer behavior. This integrated approach has consistently shown that factors such as enjoyment, perceived usefulness, and social norms significantly shape user attitudes toward technology-based commerce. Complementary evidence from related fields—particularly mobile payment systems—emphasizes the need to account for technology-specific features when analyzing adoption. Additional studies have similarly shown that these factors play a critical role in shaping user engagement with digital platforms, including voice shopping (Al-Maghrabi et al., 2011; R. Hasan et al., 2021).

This approach is further reinforced by considering elements such as performance expectancy and social influence, which are identified as key drivers in online shopping intentions. Previous research underlines the importance of these factors, pointing out that performance expectancy, or how well a user believes the technology will meet their needs, alongside the influence exerted by social circles, significantly shapes consumer decisions in the online shopping domain (Musleh et al., 2015).

By merging the insights from both TAM and UGT, this approach offers a more nuanced understanding of voice shopping. It acknowledges that while functional attributes like performance expectancy are crucial, the subjective experiences and social dimensions provided by UGT are equally vital in comprehending the full spectrum of factors that motivate consumers to engage with voice shopping technologies. This point of view is essential for capturing the complex interplay of individual preferences, social dynamics, and perceived technological utility in shaping consumer behavior in the digital commerce landscape.

The application of the Uses and Gratifications Theory (UGT) has been instrumental in understanding the motivations behind purchase intentions, focusing on variables that extend beyond the characteristics of the technology and connect more closely with individual

attitudes. Within these key gratifications, we find enjoyment, performance expectancy, and social norms as pivotal factors influencing these intentions. Various studies corroborate the role of enjoyment in shaping attitudes toward technology use, suggesting its direct impact on purchase decisions. For instance, Hasan (2015) highlights how enjoyment can significantly influence consumer decisions to buy. Additionally, the importance of performance expectancy in predicting continued use of online platforms, which is a direct indicator of purchasing behavior, is well-documented, as seen in (Agarwal & Sahu, 2021).

Moreover, the influence of social norms on consumer behavior cannot be overlooked. Previous studies emphasize the significance of social influences in shaping how consumers interact with and perceive digital platforms, underscoring the critical role of UGT gratifications in shaping consumer purchase behaviors in various digital contexts (Ham et al., 2014). By examining these aspects, UGT provides a comprehensive view of the multifaceted nature of consumer behavior in the digital era, acknowledging that individual attitudes and social context are as influential as the technological attributes in determining purchase intentions.

This blend of TAM and UGT allows us to delve into not just the functional attributes like ease of use and usefulness, but also into the gratifications driving users towards voice shopping, such as convenience and personalization.

1.3. Proposed model and hypothesis development

Our proposed model, grounded in the TAM and the UGT, aims to comprehensively examine consumer behavior in voice shopping, particularly through voice assistants like the Echo Dot. This model integrates traditional TAM variables – Perceived Usefulness, Perceived Ease of Use, and Perceived Safety – with key gratifications from UGT, namely Enjoyment, Performance Expectancy, and Social Norms, to provide a thorough understanding of consumer behavior in the context of voice shopping.

TAM plays a crucial role in understanding online consumer behavior, especially in e-commerce and voice shopping. Key TAM variables, such as Perceived Usefulness, Perceived Ease of Use, and Perceived Safety, have been shown to significantly influence consumer purchase intentions. Previous studies have emphasized the role of Perceived Usefulness in shaping intentions towards voice-activated shopping and Perceived Ease of Use

in the acceptance of technologies (Jung & Yeo, 2015; Sorensen & Jorgensen, 2021). Additionally, Perceived Safety, especially regarding information security in online shopping, has been highlighted as a significant factor (Zaharia & Würfel, 2021). Below we present the variables of the study:

1.3.1. Perceived Usefulness and Perceived Ease of Use

In the realm of voice shopping, the concept of Perceived Usefulness. It refers to the degree to which a user believes that using a system will enhance their shopping experience (Davis, 1989). Supporting this, Sorensen & Jorgensen (2021) found that perceived usefulness, alongside perceived enjoyment and innovativeness, significantly impacts millennials' behavioral intention to use voice-activated shopping technologies. In parallel, Perceived Ease of Use, another fundamental aspect of TAM, signifies the extent to which a person believes that using a system will be effort-free (Davis, 1989). This factor is essential in voice shopping, where systems that are user-friendly are more likely to gain acceptance among consumers.

1.3.2. Perceived Safety

Perceived Safety in online shopping, especially in the context of voice commerce, includes the perceived risk of information security and privacy. Consumers' perception of safety and risk significantly influences their willingness to engage in online shopping. For instance, perceived risk was found to negatively influence the intention to use smart speakers for online shopping, highlighting the importance of perceived safety in the decision-making process (Zaharia & Würfel, 2021).

1.3.3. Enjoyment

Enjoyment is crucial for the acceptance of online games, significantly affecting users' behavioral intentions (Merhi, 2016). These findings suggest that the enjoyment aspect of Uses and Gratification Theory is directly linked to perceived usefulness, as it enhances user satisfaction and positive attitudes towards technology use.

In the context of voice shopping, enjoyment refers to the pleasure and satisfaction users derive from the experience. This form of gratification is crucial as it positively

influences attitudes toward the acceptance of voice shopping technologies. Enjoyment enhances the user experience, making it more likely for individuals to perceive voice shopping as useful and beneficial (Hasan, 2015).

1.3.4. Performance Expectancy

Performance expectancy in voice shopping pertains to the anticipated improvement in shopping efficiency and effectiveness. This gratification, coming from the UGT, is a significant predictor of user satisfaction and the subsequent intention to continue using voice shopping platforms. It underscores the practical benefits and conveniences offered by voice shopping, thereby reinforcing its perceived usefulness (Agarwal & Sahu, 2021).

1.3.5. Social Norms

Social norms refer to the influence of societal expectations and peer behavior on individual actions. In the context of social media, subjective norms were found to be influential in shaping users' attitudes towards content creation, although they did not directly affect behavior (Ham et al., 2014). This indicates that while social norms can guide user perceptions of usefulness, their direct impact on user behavior might be limited.

Numerous studies have drawn attention to the significant impact of a positive state of mind on perceived usefulness, particularly within various technological contexts and consumer behaviors. Maintaining a positive mindset can profoundly influence the perception of usefulness, especially during shopping activities (Cai & Xu, 2006). Furthermore, users' perceptions of enjoyment can indeed wield influence over their perception of a technology's usefulness (Davis, 1986).

The concept of enjoyment plays a crucial role in shaping user engagement with technology, similar to its importance in the TAM3 (Venkatesh & Bala, 2008). Within UGT, enjoyment is not just a peripheral aspect but a central factor that directly influences users' gratification and satisfaction derived from technology use.

Our model, grounded in UGT and TAM, emphasizes enjoyment as a fundamental predictor of consumer acceptance of voice shopping technologies. This focus is informed by the work of Pagani (2004) and Curran & Meuter (2007), whose research underscores the

significance of enjoyment in shaping user experiences with new technologies. By adopting Pagani (2004) definition, we ensure a consistent and comprehensive approach to analyzing enjoyment within our UGT-based framework.

Thus, our hypothesis posits that in the realm of voice shopping, enjoyment significantly contributes to user satisfaction and gratification, subsequently affecting their acceptance and continued use of these technologies. This hypothesis reflects the UGT's emphasis on personal fulfillment and pleasure as key drivers of technology engagement.

H1. Enjoyment exerts a positive effect on the Perceived Usefulness of voiced virtual assistants.

In the context of the UGT, Performance Expectancy, which refers to users' belief in a technology's capability to enhance task performance, plays a pivotal role in shaping their gratifications or satisfaction derived from the technology. This concept focuses on how effectively a technology meets users' needs and desires (paralleling Davis, 1989).

Research within the UGT paradigm indicates that the greater the expectation of a technology's efficiency, the more gratification users derive from it (Venkatesh et al., 2003). This concept under UGT emphasizes that when users anticipate that a technology will efficiently meet their personal or professional goals, it enhances their satisfaction and fulfillment, leading to continued usage (Agarwal & Sahu, 2021). Therefore, in the UGT framework, Performance Expectancy is closely tied to the fulfillment of users' specific needs and desires, showing that effective performance is crucial for higher user gratification and continued engagement with the technology.

H2. Performance expectancy has a positive effect on the Perceived Usefulness of voiced virtual assistants.

Social Norms, understood as the collective expectations and behaviors within one's social circle, play a critical role in the UGT regarding an individual's perception of technology's benefits. These norms are seen as societal standards or guidelines that dictate appropriate behavior in specific social contexts (Cialdini & Trost, 1998). In the UGT framework, the impact of Social Norms on technology acceptance is attributed to the influence and pressure to conform to these prevailing norms.

Research within the UGT context highlights the influence of Social Norms in shaping users' perceptions of a technology's benefits in TAM (Venkatesh et al., 2003). When individuals sense that their social or professional network expects them to adopt a certain technology, this perception enhances their belief in the technology's benefits. Within UGT, Social Norms serve as a social driver, encouraging individuals to see a technology as more advantageous if it aligns with peer or colleague expectations. This reflects the UGT's emphasis on the social and psychological motivations behind media and technology usage.

H3. Social norm has a positive effect on the Perceived Usefulness of voiced virtual assistants.

Perceived usefulness is vital in the context of technology acceptance, particularly in the case of VAs. Perceived usefulness refers to an individual's belief about the usefulness of a technology product or service, and it has been shown to be a significant predictor of acceptance and usage behavior (Davis, 1989). Studies have found that consumers are more likely to adopt and use technology products and services that they perceive to be useful (Venkatesh & Davis, 2000). In the case of VAs, consumers may view them as useful for tasks such as setting reminders, playing music, or controlling other smart home devices (Ballati et al., 2018). Additionally, consumers may perceive VAs to be useful as a means of accessing information or completing tasks quickly and efficiently (Venkatesh & Davis, 2000).

Perceived usefulness exerts a more direct and robust influence on technology acceptance intention compared to perceived ease of use (Kramer et al., 2015). Perceived usefulness takes precedence as a primary factor in an individual's intention to adopt new technology, while perceived ease of use plays a secondary role (Davis, 1989). Perceived usefulness has positive effects on users' attitudes and behavioral intention to accept new learning systems (Gong et al., 2004). Bhattacherjee & Hikmet (2008) reached the conclusion that the perceived usefulness of information technology users has a positive impact on their intention to use that particular technology. Additionally, Pai & Huang (2011) discovered a positive association between perceived usefulness and users' intention to adopt a health information system. Rauniar et al. (2014) showed that the perceived usefulness of a new technology is positively associated with the intention to use that technology. Finally, Lunney et al. (2016) showed that perceived utility positively influences user attitudes and acceptance of smart devices.

Perceived ease of use refers to how much an individual thinks that employing a particular system would entail little effort or difficulty (Davis, 1989). Perceived ease of use has long been considered an important factor in the acceptance of technology. The original TAM indicated that perceived ease of use works through perceived usefulness and has an indirect effect on behavioral intention to use. Davis (1989) in his original TAM model, proposed relationships between what we consider to be traditional variables. We adopt his approach and propose the following hypotheses:

H4. The perceived ease of use of voiced virtual assistants is positively related to voice shopping intention.

H5. The perceived usefulness of voiced virtual assistants is positively related to voice shopping intention.

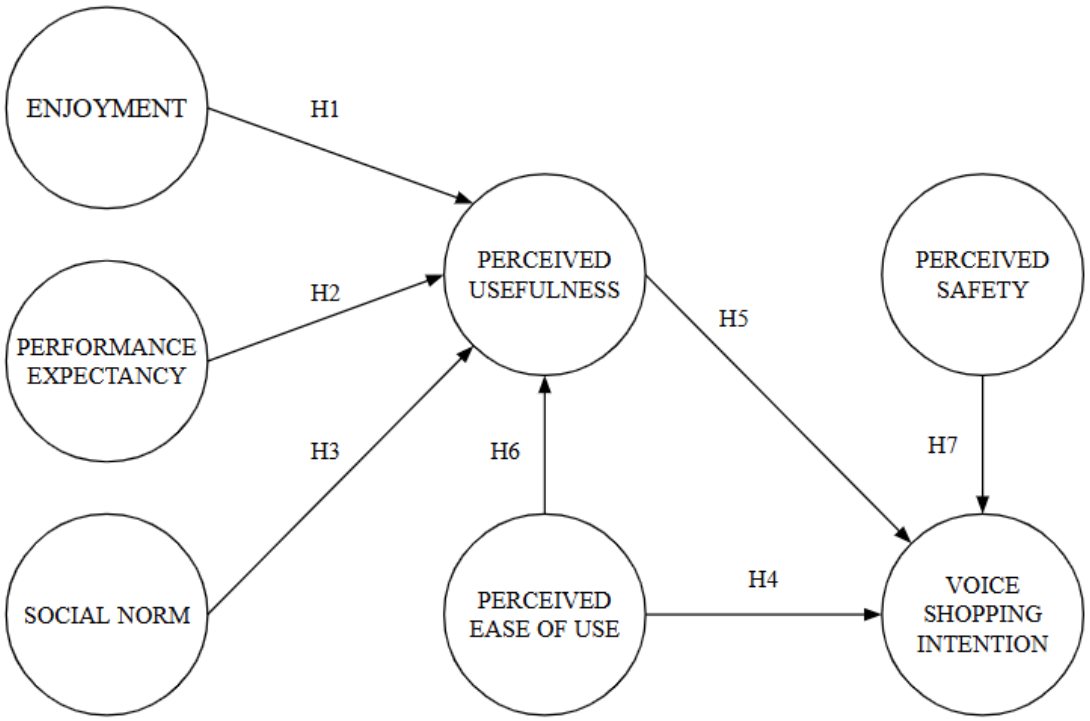
H6. The perceived ease of use of voiced virtual assistants is positively related to perceived usefulness of voiced virtual assistants.

In recent years, the definition of perceived safety has changed due to changing customer behavior and the shift to online transactions. It was originally called perceived risk and was limited to concepts such as fraud or product quality, but, currently, it is described as the sense of uncertainty about potential adverse outcomes associated with the utilization of a product or service (Featherman & Pavlou, 2003). These authors identified various types of risks and argued that it is important to include the perceived risk variable in the TAM. Customers relate to each other and assess risk when evaluating products/services for purchase, which can cause anxiety and discomfort. Various studies have concluded that perceived safety is an important factor in VA acceptance (Chang & Chen, 2014; Laforet & Li, 2005). Studies on acceptance have demonstrated that perceived risk stands out as a significant factor causing people's hesitancy towards utilizing this technology (Bawack et al., 2021a).

H7. The perceived safety of VAs is positively related to voice shopping intention.

The proposed model is shown in Figure 1.1. This model aims to offer an extensive understanding of how these factors collectively influence the acceptance and intention of using voice assistants for shopping. By assessing these variables, the model intends to provide valuable insights into the dynamics of consumer behavior in the evolving domain of voice commerce.

Figure 1.1. Proposed model



1.4. Methodology

The data collection process took place in Spain in June 2022 and involved conducting an online survey through a market research company. Participants were selected based on their familiarity with VA technology in their daily lives and the concept of voice shopping. All participants had previous experience with VAs and were aware of their existence. They accessed a dedicated web platform designed for this study, where they were required to answer questions regarding their perceptions of VAs and voice shopping.

We selected an age limit of 50 due to the pilot study results, which showed that users older than 50 are generally unfamiliar with the concept of VA. The final survey sample consisted of 906 users of VAs, with 49% being male and 51% being female. The age range of the participants varied from 18 to 50 years.

To ensure the survey's content validity in the context of VAs, an extensive review of the literature was conducted to tailor the variables within the model appropriately. The survey then underwent a rigorous validation process involving multiple experts. All the variables in the survey were evaluated using a seven-point Likert scale, where respondents could express

their agreement or disagreement, with 1 indicating strong disagreement and 7 indicating strong agreement. Before the main survey, a pilot study was carried out to confirm that respondents comprehended the survey items effectively.

Enjoyment, Performance Expectancy and Social Norm was adapted from the scales of Hedeman (2021) and each one consisted of three items. Perceived safety was adapted from the scale of McLean & Osei-Frimpong (2019) and comprised two items. Perceived usefulness was adapted from the scales of Moriuchi (2019) and consisted of three items. Perceived ease of use was adapted from the scales of Sarkar et al. (2020) and comprised two items. Lastly, the scale for measuring voice-shopping intention was derived from the work of Bawack et al. (2021b) and consisted of three items.

1.5. Results

1.5.1. Testing the measurement model

To investigate the proposed hypotheses, we utilized Structural Equation Modeling (SEM), an advanced statistical method that combines multiple regression with confirmatory factor analysis to simultaneously estimate a series of interrelated dependency relationships. In recent years, there has been a growing number of studies in the field of information systems employing SEM to test proposed models (Sarstedt et al., 2022).

To ensure the reliability, dimensionality, and validity of our measurement scales, we conducted both exploratory and confirmatory factor analyses using statistical software including SPSS (version 24) and Smart PLS 4. Internal reliability was assessed using Cronbach's alpha for all variables, which exceeded the required minimum value of 0.70, indicating that our variables exhibited satisfactory levels of reliability (Henseler et al., 2016). Convergent validity was confirmed by checking item loadings, all of which exceeded the recommended threshold of 0.70 (Hair, 2009).

The composite reliability values for all variables exceeded 0.70, signifying appropriate internal consistency of the variables (Aguirre-Urreta et al., 2013). Additionally, the average variance extracted (AVE) demonstrated consistency with values equal to or greater than 0.50, indicating that each variable captured a sufficient amount of variance relative to measurement errors (Fornell & Larcker, 1981). In summary, our measurement scales met the criteria for reliability. The results of these analyses are presented in Table 1.2.

Table 1.2. Measurement model results

Variable	Item	Mean	SD	λ	t-value	CA	CR	AVE
ENJ	ENJ1	3.56	1.724	0.931	47.329***	0.842	0.842	0.863
	ENJ2	3.89	1.830	0.928	52.102***			
PEXP	PEXP1	3.37	1.632	0.937	50.390***	0.823	0.847	0.848
	PEXP2	3.28	1.791	0.904	54.965***			
SNOR	SNOR1	2.59	1.781	0.934	50.772***	0.870	0.876	0.885
	SNOR2	3.09	1.738	0.947	47.394***			
PU	PU1	3.86	1.682	0.939	65.026***	0.845	0.928	0.865
	PU2	4.35	1.728	0.921	82.707***			
PEOU	PEOU1	5.74	1.215	0.912	25.698***	0.785	0.786	0.823
	PEOU2	5.39	1.251	0.902	26.895***			
PSAF	PSAF1	4.18	1.695	0.970	81.223***	0.938	0.938	0.941
	PSAF2	4.16	1.666	0.970	81.920***			
VSHOP	VSHOP1	5.09	1.602	0.833	33.199***	0.729	0.724	0.650
	VSHOP2	5.18	1.605	0.850	40.420***			
	VSHOP3	3.35	1.756	0.730	25.019***			

Notes: CA: Cronbach's alpha; CR: composite reliability; AVE: average variance extracted.
***sig. < 0.01.

ENJ: Enjoyment; PEXP: Performance Expectancy; SNOR: Social Norm; PU: Perceived Usefulness; PEOU: Perceived Ease of Use; VSHOP: Voice Shopping Intention

We placed significant importance on establishing discriminant validity, a crucial aspect in confirming that the measurement scales used in our study effectively represent separate and distinct concepts. To accomplish this, we relied on the AVE values (as displayed in Table 1.3) and conducted a comprehensive evaluation of the squared correlation coefficients found in the respective rows and columns (as presented in Table 1.3). This meticulous examination adhered closely to the established criteria recommended by (Fornell & Larcker, 1981; Hair, 2009). Our efforts in this regard provided robust evidence supporting

the presence of discriminant validity.

Table 1.3 Correlations among factors

	VSHOP	ENJ	PEXP	SNOR	PEOU	PU	PSAF
VSHOP	1,000	0,535	0,547	0,472	0,431	0,643	0,624
ENJ	0,535	1,000	0,757	0,704	0,288	0,662	0,501
PEXP	0,547	0,757	1,000	0,830	0,226	0,709	0,490
SNOR	0,472	0,704	0,830	1,000	0,123	0,605	0,443
PEOU	0,431	0,288	0,226	0,123	1,000	0,324	0,404
PU	0,643	0,662	0,709	0,605	0,324	1,000	0,493
PSAF	0,624	0,501	0,490	0,443	0,404	0,493	1,000

ENJ: Enjoyment; PEXP: Performance Expectancy; SNOR: Social Norm; PU: Perceived Usefulness; PEOU: Perceived Ease of Use; VSHOP: Voice Shopping Intention

Furthermore, we conducted a thorough examination of the HTMT (Heterotrait-Monotrait) values, as advised by Henseler et al. (2016). These values play a crucial role in assessing the adequacy of the measurement model, with HTMT values below 0.95 typically deemed appropriate, particularly for concepts that exhibit a high degree of similarity. Our comprehensive evaluation, as depicted in Table 1.4, demonstrated that all HTMT values remained below the 0.95 threshold, providing further confirmation of the strength and validity of our measurement model.

Table 1.4. Heterotrait-monotrait ratio (HTMT)

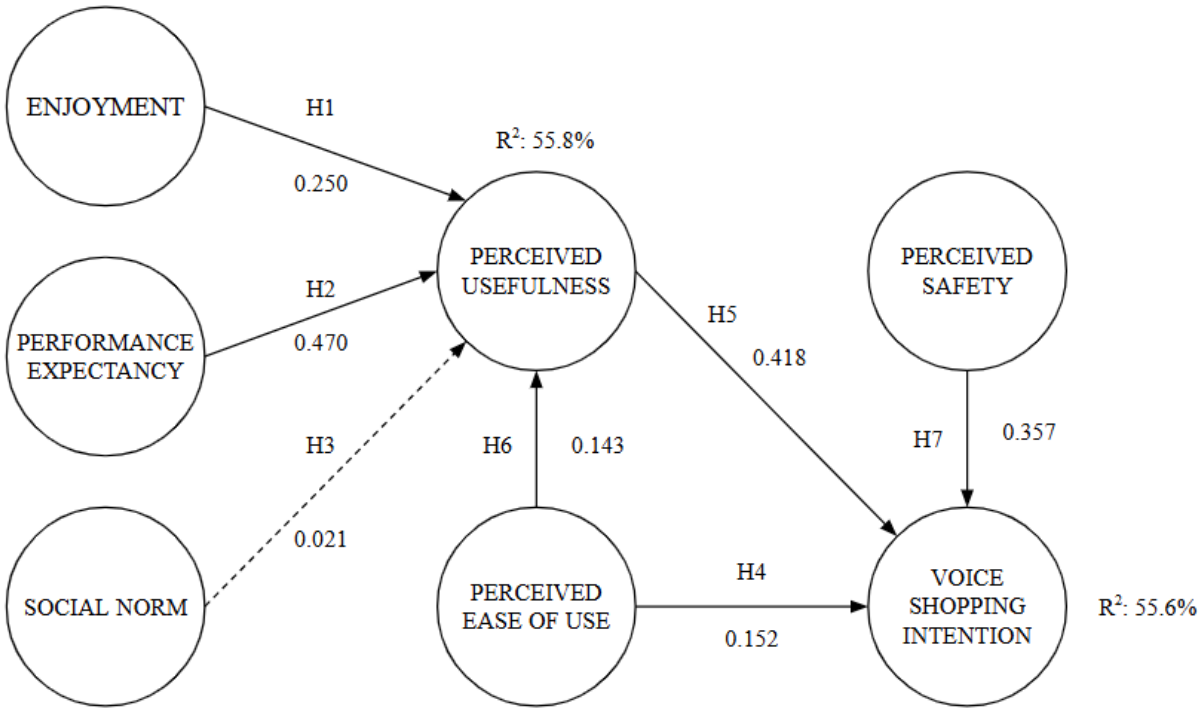
	VSHOP	ENJ	PEXP	SNOR	PEOU	PU	PSAF
VSHOP							
ENJ	0,659						
PEXP	0,671	0,902					
SNOR	0,561	0,821	0,985				
PEOU	0,579	0,355	0,276	0,147			
PU	0,799	0,782	0,838	0,700	0,398		
PSAF	0,750	0,564	0,555	0,489	0,471	0,553	

ENJ: Enjoyment; PEXP: Performance Expectancy; SNOR: Social Norm; PU: Perceived Usefulness; PEOU: Perceived Ease of Use; VSHOP: Voice Shopping Intention

1.5.2. Testing the structural model

We evaluated the structural model's path analysis using a bootstrapping method with 5,000 re-samples. The current structural model explained 52.6% of the voice shopping variance. The results indicate a satisfactory model as all of the structural model's variance percentages explained were higher than 10% (Falk & Miller, 1992).

Figure 1.2. Estimated model



Note: Significance of all relationships $p < 0.01$.

The dashed line (------) indicates a non-significant relationship.

Most of the hypotheses received support in the study (as indicated in Table 1.5). The research findings demonstrate a significant and positive relationship between Enjoyment and Perceived Usefulness (H1) as well as between Performance Expectancy and Perceived Usefulness (H2). These results align with the propositions made by Cai & Xu (2006) and Venkatesh et al. (2003). However, the relationship between Social Norm and Perceived Usefulness (H3) did not yield significant results.

Furthermore, the association between Perceived Ease of Use and Perceived Usefulness was found to be significantly positive (H4). Regarding the hypotheses related to Behavioral Intention, all of them exhibited statistical significance. H5, H6, and H7, which

examined the relationships involving Perceived Usefulness (H5), Perceived Ease of Use (H6), and Perceived Safety (H7), were all supported.

Table 1.5. Proposed hypotheses

Hypothesis	Dependent variable	Proposed effect	Result
H1	Perceived Usefulness	Higher with Enjoyment	Supported
H2	Perceived Usefulness	Higher with Performance Expectancy	Supported
H3	Perceived Usefulness	Higher with Social Norm	Not supported
H4	Perceived Usefulness	Higher with Perceived Ease of Use	Supported
H5	Voice Shopping Intention	Higher with Perceived Usefulness	Supported
H6	Voice Shopping Intention	Higher with Perceived Ease of Use	Supported
H7	Voice Shopping Intention	Higher with Perceived Safety	Supported

Table 1.6 captures the indirect effect of gratifications from the model on the variable 'voice shopping intention'. We can verify the indirect effect between enjoyment and voice shopping intention through perceived usefulness, with a significant t-value of 4.928 and a p-value of 0.000. Similarly, there is an indirect effect between Performance Expectancy and Voice Shopping Intention, evidenced by a t-value of 7.915 and a p-value of 0.000. An indirect relationship occurs when one or more intervening variables are present, serving to transmit the influence of variable X on variable Y, as explained by Baron & Kenny (1986). This is particularly relevant in mediation analysis, where researchers explore whether changes in an independent variable lead to changes in a mediating variable, which in turn lead to changes in the dependent variable.

Table 1.6. Indirect effects

	Standard deviation	T statistics	P values
ENJ -> VSHOP	0.021	4.928	0.000
PEXP -> VSHOP	0.025	7.915	0.000

ENJ: Enjoyment; VSHOP: Voice Shopping Intention; PEXP: Performance Expectancy

1.6. Conclusions and discussion

Our study had a central objective: to develop a theoretical framework capable of explaining and potentially predicting the acceptance voice shopping intention. Specifically, we focused on extending a TAM model with various gratifications obtained from the UGT. Previous research has separately analyzed the acceptance of virtual assistants and voice shopping. However, the simultaneous examination of both topics has not yielded significant conclusions. In contrast to other published studies that primarily focus on perceived risks associated with such technologies, our work emphasizes the utility of this new technology in voice commerce. Our research aimed to elucidate the role of utility in shaping users' perceptions during voice shopping interactions. The findings of our study revealed a significant impact of performance expectancy and utility perceived on behavioral intention, indicating a favorable attitude among users toward voice shopping.

The utilization of gratifications allows for a more comprehensive explanation of the intention to use VAs for voice shopping when compared to the use of simpler models of technology acceptance such as TAM and UTAUT. Our study demonstrates that the presence of enjoyment, performance expectation, and perceived safety, as gratifications, contributes to explaining the acceptance of voice shopping with VAs as a phenomenon.

As anticipated, the results provided support for most of the hypotheses pertaining to gratifications. The outcomes presented in Table 7 and Figure 2 underscore the significance of enjoyment and performance expectancy concerning perceived usefulness (H1 and H2). The presence of features related to enjoyment and performance expectancy contributes to an explanation of the perception of usefulness in voice shopping with VAs. These findings align with McLean & Osei-Frimpong (2019) where the authors also highlighted the importance of emotions and expectations when using VAs, as evidenced through gratifications and emphasizing the role of emotions in shaping user experiences and expectations with VAs, and focusing on how user expectations and the gratifications sought from VAs influence their overall satisfaction and continued usage.

However, Hypothesis 3, which concerns the relationship between Social Norm and Perceived Usefulness, does not follow the same pattern. This phenomenon can be understood through the lens of the Diffusion of Innovation Theory (Rogers et al., 2014), which suggests that Social Norms can impact technology acceptance decisions when there is a perceived

alignment between the innovation and existing social norms. If users believe that using VAs for voice shopping is not in harmony with established social norms, or if these norms do not support this technology, then the influence of social norms on technology adoption may be weakened.

Furthermore, the Diffusion of Innovation Theory proposes that the adoption of an innovation is influenced by how much this innovation aligns with each user's existing experiences and values. The absence of a significant relationship between social norms and perceived usefulness could be due to a perceived misalignment between these two factors. In simpler terms, this insignificant relationship might be attributed to a perceived lack of compatibility between societal expectations and the practical utility of using VAs for voice shopping. This could be because users already have prior experience with many technologies, and they might no longer expect to be socially accepted solely based on their use of technology.

While the results regarding the relationships between PEOU and PU (H4) and PEOU and VSHOP (H6) are statistically significant, it's essential to highlight the limited effect that the PEOU variable has on these relationships. Some studies have suggested that the influence of perceived ease of use on technology acceptance may have weakened over time, especially in the context of adopting VAs (Venkatesh, 2000; Venkatesh & Davis, 1996; Wu & Wang, 2005; Yang & Lee, 2019), through pre-implementation testing, demonstrated that perceived ease of use has a direct and significant impact on an individual's intention to use a system before they have any direct experience with it. However, as individuals gain experience with the system over time, the influence of perceived ease of use on their intention to use gradually diminishes and eventually becomes negligible. Although the direct influence of perceived ease of use on behavioral intention to use is weak, it does exert an indirect impact on behavioral intention through perceived usefulness, as concluded by Wu & Wang (2005).

The results of H5 and H7 highlight the significance of perceived usefulness and perceived safety, respectively, when it comes to adopting new technology. Therefore, a higher level of perceived usefulness (preceded by enjoyment and performance expectancy) is essential for facilitating the acceptance of voice shopping. Similarly, a higher degree of perceived safety is also necessary for the acceptance of voice shopping with VAs. These findings align with recent and prior literature (Calahorra-Candao & Martín-de Hoyos, 2024),

suggesting that one of the primary factors influencing the acceptance of this technology is the confidence users have in the device's security.

1.6.1. Implications for theory

We believe this study provides a valuable contribution to the relatively limited existing literature on voice shopping with virtual assistants. There is a scarcity of studies that specifically apply TAM and UGT to VAs, and the available research may be somewhat biased. This study has the potential to enrich the literature by applying TAM theory to VAs in the underexplored domain of voice shopping. Given the unique characteristics of VAs, many users may not fully comprehend their utility, leading to unknown conclusions. Therefore, we found it essential to investigate the factors that could enhance the behavioral intention of using VAs for voice shopping.

Furthermore, this study significantly enhances the literature on voice shopping with virtual assistants by identifying two key gratifications in the acceptance of this technology: perceived usefulness and perceived safety. The influence of the former can be partly attributed to the emotions that arise from using these technologies and the expectations associated with them. The presence of the second variable finds support in prior literature, emphasizing the users' requirement for perceiving a technology as secure before embracing it (Calahorra-Candao & Martín-de Hoyos, 2024).

Unlike other studies that have focused on specific VAs, this study has deliberately avoided such bias by not concentrating on any particular virtual assistant. Prior publications, as exemplified by Ewers et al. (2020), often narrow their focus to a particular existing virtual assistant, which could be influenced by users' prior judgments and knowledge. Despite the growing interest in VAs in the literature, there is limited understanding of the role of voice shopping using these devices. Currently, there is no consensus on the factors influencing the acceptance of this phenomenon, especially those associated with the characteristics of VAs.

This study served two primary purposes. Firstly, it allowed us to contribute to the literature on virtual assistants by proposing a different perspective on their acceptance. Secondly, it shed some light on the phenomenon of voice shopping and the factors that may be necessary for its acceptance.

1.6.2. Implications for practice

This study holds significant practical implications across various fields including AI-related research, development of voice-based software, and advertisement strategies. Echoing the findings of Davis (1989) in the initial TAM, this research underscores that developers of VAs have successfully instilled in users a sense of usefulness for the services offered by VAs, thereby fostering user interaction with these systems. It highlights that while developing functional aspects of devices is essential, prioritizing the emotional component, such as user enjoyment, is becoming increasingly vital.

Moreover, the study emphasizes the critical importance of security and privacy in the acceptance of new AI technologies. The ability to effectively communicate and ensure the security of transactions and the confidentiality of personal data is identified as a key determinant for user trust and acceptance. For companies in this space, it's crucial to manage and communicate realistic expectations to avoid creating false hopes or misconceptions among users, considering the significant role that these expectations play in the adoption of new technologies. This comprehensive approach, which balances functional, emotional, and security aspects, is essential for the successful integration and acceptance of AI advancements in everyday technology use.

Taking into account the numerous daily interactions that occur between users through various platforms, providing information about the benefits of using VAs for voice purchases is crucial for their development. However, as mentioned at the beginning of this study, the use of VAs for voice purchases is still limited. It should be noted that, for e-commerce, this technology provides consumers with the ability to use voice commands to make purchases. Therefore, based on the results obtained in this study, the emotional connection that develops between the user and the VA through a human feature, such as a human voice, will influence key variables such as enjoyment during the shopping experience, perceived safety, and perceived usefulness of voice-assisted purchases. The current challenge is to find the right voice that can connect with the user and convey emotions of safety, fun, usefulness, and ease of the purchase process, ultimately creating a journey of positive emotions for the user.

1.6.3. Limitations and future research

Like many studies, this research has its limitations. The study focused on general

information and purchase searches, and it would be valuable to test the empirical model for specific product or information searches. It is reasonable to assume that the significance of localization may differ depending on the type of search task conducted, such as hedonic versus utilitarian product searches.

This study did not consider the various VA devices available in the market, each of which has a unique voice. Future research could involve examining VAs to determine which has better acceptance for voice shopping and which voice has the most and least effects. In addition, future researchers can enhance their understanding of the influence of VAs on consumer engagement by incorporating additional anthropomorphic factors into the model. Lastly, we see the in-depth study of the emotions transmitted by these devices as essential, as suggested by the study by Huh et al. (2023), which makes evident the need to understand the role of emotions in order to understand the user experience. Analyzing what emotions are transmitted in a possible shopping scenario could shed light on what the necessary emotional conditions should be to facilitate and encourage voice shopping.

1.7. References

Adolphs, E., & Zaharia, S. (2021). Consumers' Acceptance of a Voice Commerce Application in FMCG in Germany, US and UK. *International Conference on Human-Computer Interaction*, 3-21.

Agarwal, V., & Sahu, R. (2021). Predicting Repeat Usage Intention Towards O2O Food Delivery: Extending UTAUT2 with User Gratifications and Bandwagoning. *Journal of Foodservice Business Research*, *25*(5), 434-474. https://doi.org/10.1080/15378020.2021.1951074

Aguirre-Urreta, M. I., Marakas, G. M., & Ellis, M. E. (2013). Measurement of composite reliability in research using partial least squares: Some issues and an alternative approach. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, *44*(4), 11-43. https://doi.org/10.1145/2544415.2544417

Ajzen, I. (1980). *Understanding Attitudes and Predicting Social Behavior*. Prentice-Hall.

Al-Maghrabi, T., Dennis, C., Halliday, S. V., & BinAli, A. (2011). Determinants of Customer Continuance Intention of Online Shopping. *International Journal of Business Science & Applied Management (IJBSAM)*, *6*(1), 41-66.

Amazon Staff. (2022, noviembre 30). *Amazon announces its biggest holiday shopping weekend ever: Top sellers included the Nintendo Switch and the Echo Dot*. US About Amazon. https://www.aboutamazon.com/news/retail/amazon-best-sellers-black-friday-cyber-monday

Astound Commerce. (2023). *Change in smart speaker purchases by region 2021*. Statista. https://www.statista.com/statistics/1276992/change-shopping-via-smart-speaker-worldwide-region/

Bagozzi, R. P. (2007). The Legacy of the Technology Acceptance Model and a Proposal for a

Paradigm Shift. *Journal of the Association for Information Systems*, *8*(4), 3.

Ballati, F., Corno, F., & De Russis, L. (2018). "Hey Siri, Do You Understand Me?": Virtual Assistants and Dysarthria. En *Intelligent Environments 2018* (pp. 557-566). IOS Press. https://doi.org/10.3233/978-1-61499-874-7-557

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173-1182. https://doi.org/10.1037/0022-3514.51.6.1173

Bawack, R. E., Wamba, S. F., & Carillo, K. D. A. (2021a). Adoption of Smart Speakers for Voice Shopping. En F. Ceci, A. Prencipe, & P. Spagnoletti (Eds.), *Exploring Innovation in a Digital World* (pp. 21-35). Springer International Publishing. https://doi.org/10.1007/978-3-030-87842-9_3

Bawack, R. E., Wamba, S. F., & Carillo, K. D. A. (2021b). Exploring the role of personality, trust, and privacy in customer experience performance during voice shopping: Evidence from SEM and fuzzy set qualitative comparative analysis. *International Journal of Information Management*, *58*, 102309. https://doi.org/10.1016/j.ijinfomgt.2021.102309

Bhattacherjee, A., & Hikmet, N. (2008). Reconceptualizing Organizational Support and its Effect on Information Technology Usage: Evidence from the Health Care Sector. *Journal of Computer Information Systems*, *48*(4), 69-76. https://doi.org/10.1080/08874417.2008.11646036

Cai, S., & Xu, Y. (2006). Effects of outcome, process and shopping enjoyment on online consumer behaviour. *Electronic Commerce Research and Applications*, *5*(4), 272-281. https://doi.org/10.1016/j.elerap.2006.04.004

Calahorra-Candao, G., & Martín-de Hoyos, M. J. (2024). The effect of anthropomorphism of

virtual voice assistants on perceived safety as an antecedent to voice shopping. *Computers in Human Behavior*, *153*, 108124.

Celik, H. (2016). Customer Online Shopping Anxiety within the Unified Theory of Acceptance and Use Technology (UTAUT) Framework. *Asia Pacific Journal of Marketing and Logistics*, *28*(2).

Chang, C.-H., & Chen, Y.-S. (2014). Managing green brand equity: The perspective of perceived risk theory. *Quality & Quantity*, *48*(3), 1753-1768. https://doi.org/10.1007/s11135-013-9872-y

Cheng, Y., & Jiang, H. (2020). How do AI-driven Chatbots Impact User Experience? Examining Gratifications, Perceived Privacy Risk, Satisfaction, Loyalty, and Continued Use. *Journal of Broadcasting & Electronic Media*, *64*(4), 592-614.

Chung, D., Kim, H., & Ahn, S. (2022). An Integrated Study of User Acceptance and Resistance on Voice Commerce. *International Journal of Innovation and Technology Management*, *19*(07), 2250025.

Cialdini, R. B., & Trost, M. R. (1998). Social Influence: Social Norms, Conformity, and Compliance. En *The Handbook of Social Psychology* (Vol. 2, pp. 151-192). McGraw-Hill.

Curran, J. M., & Meuter, M. L. (2007). Encouraging Existing Customers to Switch to Self-Service Technologies: Put a Little Fun in their Lives. *Journal of Marketing Theory and Practice*, *15*(4), 283-298. https://doi.org/10.2753/MTP1069-6679150401

Davis, F. D. (1986). A technology acceptance model for empirically testing new end-user information systems: Theory and results. *Ph. D. dissertation, Massachusetts Institute of Technology*. https://cir.nii.ac.jp/crid/1570291224355974272

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13*(3), 319-340.

https://doi.org/10.2307/249008

Dousay, T. A., & Hall, C. (2018). *"Alexa, tell me about using a virtual assistant in the classroom"*. 1413-1419. https://www.learntechlib.org/primary/p/184359/

Ewers, K., Baier, D., & Höhn, N. (2020). Siri, Do I Like You? Digital Voice Assistants and Their Acceptance by Consumers. *SMR-Journal of Service Management Research*, *4*(1), 52-68.

Falk, R. F., & Miller, N. B. (1992). *A primer for soft modeling* (pp. xiv, 103). University of Akron Press.

Featherman, M. S., & Pavlou, P. A. (2003). Predicting e-services adoption: A perceived risk facets perspective. *International Journal of Human-Computer Studies*, *59*(4), 451-474. https://doi.org/10.1016/S1071-5819(03)00111-3

Fornell, C., & Larcker, D. F. (1981). Structural Equation Models with Unobservable Variables and Measurement Error: Algebra and Statistics. *Journal of Marketing Research*, *18*(3), 382-388. https://doi.org/10.2307/3150980

Gong, M., Xu, Y., & Yu, Y. (2004). An Enhanced Technology Acceptance Model for Web-Based Learning. *Journal of Information Systems Education*, *15*(4), 365-374.

Gorgijevski, A., Lind, C. H., & Lagerström, K. (2019). Does Proactivity Matter? The Importance of Initiative Selling Tactics for Headquarters Acceptance of Subsidiary Initiatives. *Journal of International Management*, *25*, 100673.

Hair, J. (2009). Multivariate Data Analysis. *Faculty and Research Publications*. https://digitalcommons.kennesaw.edu/facpubs/2925

Ham, C. D., Lee, J., & Lee, H. S. (2014). Understanding Consumers' Creating Behaviour in Social Media: An Application of Uses and Gratifications and the Theory of Reasoned Action. *International Journal of Internet Marketing and Advertising*, *8*(4), 241-263.

Hasan, B. (2015). Applying Uses and Gratifications Theory to Acceptance of ERP Systems.

Int'l Conference on Business, Marketing & Information System Management (BMISM'15), 36-41. https://doi.org/10.15242/icehm.ed1115028

Hasan, R., Shams, R., & Rahman, M. (2021). Consumer trust and perceived risk for voice-controlled artificial intelligence: The case of Siri. *Journal of Business Research*, *131*, 591-597. https://doi.org/10.1016/j.jbusres.2020.12.012

Hedeman, J. (2021, enero). *Hey Siri, let's go shopping! : A study into the factors influencing Dutch consumers' intention to use a voice assistant for online shopping* [Info:eu-repo/semantics/masterThesis]. University of Twente. http://essay.utwente.nl/85619/

Henseler, J., Hubona, G., & Ray, P. A. (2016). Using PLS path modeling in new technology research: Updated guidelines. *Industrial Management & Data Systems*, *116*(1), 2-20. https://doi.org/10.1108/IMDS-09-2015-0382

Huang, Z., & Benyoucef, M. (2013). From e-commerce to social commerce: A close look at design features. *Electronic Commerce Research and Applications*, *12*(4), 246-259. https://doi.org/10.1016/j.elerap.2012.12.003

Huh, J., & Kim, H. Y. (2022). The Uses and Gratifications Model of Voice Shopping. *International Textile and Apparel Association Annual Conference Proceedings*, *78*(1).

Huh, J., Whang, C., & Kim, H.-Y. (2023). Building trust with voice assistants for apparel shopping: The effects of social role and user autonomy. *Journal of Global Fashion Marketing*, *14*(1), 5-19.

Jung, J., & Yeo, H. (2015). Digital Coupon Gift-giving Model through Gift-Giving Motivation. *Journal of The Korean Chemical Society*, *6*, 105-110. https://doi.org/10.15207/JKCS.2015.6.6.105

Katz, E., Blumler, J. G., & Gurevitch, M. (1973). Uses and Gratifications Research. *The Public Opinion Quarterly*, *37*(4), 509-523.

Kramer, B., DaDalt, O., Burstein, A., D'Ambrosio, L., Wahl, H., & Coughlin, J. (2015). Technology Acceptance in Dementia Care: The Impact of Perceived Usefulness and Perceived Ease of Use. *GERONTOLOGIST*, *55*, 473-473.

Kun, C., & Xing, L. (2016). The Inspection and Integration of Employee Voice Research in View of Its Nature and Effects. *Foreign Economics & Management*, *38*(08), 95-112.

Laforet, S., & Li, X. (2005). Consumers' attitudes towards online and mobile banking in China. *International Journal of Bank Marketing*, *23*(5), 362-380. https://doi.org/10.1108/02652320510629250

Liébana-Cabanillas, F., Muñoz-Leiva, F., & Sánchez-Fernández, J. (2013). The impact of risk on the technological acceptance of mobile payment services. *Global Business Perspectives*, *1*, 309-328.

Loup Ventures. (2023). *Smart speaker global shipments 2014-2025*. Statista. https://www.statista.com/statistics/1022809/worldwide-smart-speaker-unit-shipment/

Lunney, A., Cunningham, N. R., & Eastin, M. S. (2016). Wearable fitness technology: A structural investigation into acceptance and perceived fitness outcomes. *Computers in Human Behavior*, *65*, 114-120. https://doi.org/10.1016/j.chb.2016.08.007

Madlberger, M., & O'Neill, J. (2020). Raising the Voice: Exploring Antecedents of Voice-Activated Assistant Usage Intention. *Proceedings of the 19th International Conference on WWW/Internet*.

McLean, G., & Osei-Frimpong, K. (2019). Hey Alexa … examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in Human Behavior*, *99*, 28-37. https://doi.org/10.1016/j.chb.2019.05.009

Merhi, M. I. (2016). Towards a Framework for Online Game Adoption. *Computers in Human Behavior*, *60*, 253-263.

Mordor Intelligence. (2021). *Intelligent Virtual Assistant (IVA) Market—Growth, Trends,*

Covid-19 impact, and forecasts (2021-2026). https://www.mordorintelligence.com/industry-reports/intelligent-virtual-assistant-market

Moriuchi, E. (2019). Okay, Google!: An empirical study on voice assistants on consumer engagement and loyalty. *Psychology & Marketing*, *36*(5), 489-501. https://doi.org/10.1002/mar.21192

Musleh, J. S., Marthandan, G., & Aziz, N. (2015). An Extension of UTAUT Model for Palestine E–commerce. *International Journal of Electronic Business*, *12*(1), 95-115.

Naqvi, A. M. H., Jiang, Y., Miao, M., & Naqvi, M. H. (2020). The Effect of Social Influence, Trust, and Entertainment Value on Social Media Use: Evidence from Pakistan. *Cogent Business & Management*, *7*(1), 1723825. https://doi.org/10.1080/23311975.2020.1723825

Natarajan, T., Balasubramanian, S. A., & Kasilingam, D. L. (2017). Understanding the Intention to Use Mobile Shopping Applications and Its Influence on Price Sensitivity. *Journal of Retailing and Consumer Services*, *37*, 8-22.

Pagani, M. (2004). Determinants of adoption of third generation mobile multimedia services. *Journal of Interactive Marketing*, *18*(3), 46-59. https://doi.org/10.1002/dir.20011

Pai, F.-Y., & Huang, K.-I. (2011). Applying the Technology Acceptance Model to the introduction of healthcare information systems. *Technological Forecasting and Social Change*, *78*(4), 650-660. https://doi.org/10.1016/j.techfore.2010.11.007

Park, N. (2010). Adoption and Use of Computer-Based Voice Over Internet Protocol Phone Service: Toward an Integrated Model. *Journal of Communication*, *60*, 40-72. https://doi.org/10.1111/J.1460-2466.2009.01440.X

Park, S., Park, S., & Whang, M. (2022). Empathic Responses of Behavioral-Synchronization in Human-Agent Interaction. *Computers, Materials & Continua*, *71*(2).

Rauniar, R., Rawski, G., Yang, J., & Johnson, B. (2014). Technology acceptance model (TAM) and social media usage: An empirical study on Facebook. *Journal of Enterprise Information Management*, *27*(1), 6-30. https://doi.org/10.1108/JEIM-04-2012-0011

Rhee, C. E., & Choi, J. (2020). Effects of personalization and social role in voice shopping: An experimental study on product recommendation by a conversational voice agent. *Computers in Human Behavior*, *109*, 106359. https://doi.org/10.1016/j.chb.2020.106359

Rogers, E. M., Singhal, A., & Quinlan, M. M. (2014). Diffusion of Innovations. En *An Integrated Approach to Communication Theory and Research* (pp. 432-448). Routledge.

Sarkar, S., Chauhan, S., & Khare, A. (2020). A meta-analysis of antecedents and consequences of trust in mobile commerce. *International Journal of Information Management*, *50*, 286-301. https://doi.org/10.1016/j.ijinfomgt.2019.08.008

Sarstedt, M., Hair, J. F., Pick, M., Liengaard, B. D., Radomir, L., & Ringle, C. M. (2022). Progress in partial least squares structural equation modeling use in marketing research in the last decade. *Psychology & Marketing*, *39*(5), 1035-1064. https://doi.org/10.1002/mar.21640

Simms, K. (2019). How voice assistants could change the way we shop. *Harvard Business Review*, *may*, 2-7.

Sorensen, K., & Jorgensen, J. (2021). «Hey Alexa, Let's Shop»: Millennials' Acceptance of Voice-Activated Shopping. *International Journal of E-Services and Mobile Applications (IJESMA)*, *13*(1), 1-14. https://doi.org/10.4018/ijesma.2021010101

Statista Staff. (2022). *Number of voice assistants in use worldwide 2019-2024*. Statista. https://www.statista.com/statistics/973815/worldwide-digital-voice-assistant-in-use/

Sun, C., Shi, Z. (June), Liu, X., Ghose, A., Li, X., & Xiong, F. (2019). *The Effect of Voice AI on Consumer Purchase and Search Behavior* (SSRN Scholarly Paper 3480877). Social Science Research Network. https://doi.org/10.2139/ssrn.3480877

Venkatesh, V. (2000). Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model. *Information Systems Research*, *11*(4), 342-365. https://doi.org/10.1287/isre.11.4.342.11872

Venkatesh, V., & Bala, H. (2008). Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences*, *39*(2), 273-315.

Venkatesh, V., & Davis, F. D. (1996). A Model of the Antecedents of Perceived Ease of Use: Development and Test*. *Decision Sciences*, *27*(3), 451-481. https://doi.org/10.1111/j.1540-5915.1996.tb00860.x

Venkatesh, V., & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science*, *46*(2), 186-204. https://doi.org/10.1287/mnsc.46.2.186.11926

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *Vol. 27*(No. 3), 425-478.

Whang, C., & Song, S. (2019). The Impact of Perceived Humanlikeness of Voice Assistants on Purchase Intention. *International Textile and Apparel Association Annual Conference Proceedings*, *76*(1).

Wu, J.-H., & Wang, S.-C. (2005). What drives mobile commerce?: An empirical evaluation of the revised technology acceptance model. *Information & Management*, *42*(5), 719-729. https://doi.org/10.1016/j.im.2004.07.001

Yang, H., & Lee, H. (2019). Understanding user behavior of virtual personal assistant devices. *Information Systems and E-Business Management*, *17*(1), 65-87.

https://doi.org/10.1007/s10257-018-0375-1

Zaharia, S., & Würfel, M. (2021). Voice Commerce-Studying the Acceptance of Smart Speakers. *Proceedings of the 3rd International Conference on Human Interaction and Emerging Technologies: Future Applications (IHIET 2020)*, 449-454.

Zarmpou, T., Saprikis, V., Markos, A., & Vlachopoulou, M. (2012). Modeling Users' Acceptance of Mobile Services. *Electronic Commerce Research*, *12*, 225-248.1

1.8. Appendix

Table 1.7. Scales

Item	Scales
Enjoyment - Adapted from Hedeman (2021)	
ENJ1	I think virtual assistants are an exciting technology.
ENJ2	Shopping with virtual assistants can be a fun experience
Performance Expectancy - Adapted from Hedeman (2021)	
PEXP1	I believe that using virtual assistants is useful for online shopping
PEXP2	I think that using virtual assistants for online shopping is trendy
Social Norm - Adapted from Hedeman (2021)	
SOCN1	People in my close circle used virtual assistants for online shopping
SOCN2	I have heard positive opinions about online shopping with virtual assistants
Perceived Usefulness - Adapted from Moriuchi (2019)	
PU1	Using a VA would make it easier to find information about my purchases on the internet
PU2	Using a VA would save me time when looking for information about a product or service
Perceived Ease of Use - Adapted from Sarkar et al. (2020)	
PEOU1	VAs are easy to use
PEOU2	VA technology is intuitive
Perceived Safety - Adapted from McLean & Osei-Frimpong (2019)	
PSAF1	I feel safe using VA
PSAF2	I feel that VAs are a safe technology
Voice Shopping Intention - Adapted from Bawack et al. (2021b)	
VSI1	In the future I will continue to use VA in my daily life (eg.: music, weather, shopping…)

VSI2	In the future I will continue to use VA to search for purchasing information
VSI3	In the future I will use VA to shop online

STUDY 2

Perceived Safety and the Humanization of Voice Assistants

Study 2 is an extension of the work: Calahorra-Candao, G., & Martín-de Hoyos, M. J. (2024). The effect of anthropomorphism of virtual voice assistants on perceived safety as an antecedent to voice shopping. Computers in Human Behavior, 153, 108124.

2.1. Introduction

The initial purpose of virtual assistants (VAs) was to serve as a voice-based interactive function enabling consumers to utilize their smartphones in an innovative manner. Current digital assistants are being integrated into a wide range of consumer devices, including speakers, vehicles, televisions, and wearables (Fowler, 2018; Vimalkumar et al., 2021). The conversational interface of a VA stands out as one of its most notable advantages. By engaging in natural language interactions, it offers a more intuitive and user-friendly alternative to web and mobile interfaces reliant on manual input via keyboards (Zhong & Yang, 2018). Leveraging advanced artificial intelligence (AI) technologies, VAs are capable of dynamically observing and analyzing complex data in real time to facilitate information retrieval based on user queries. Furthermore, they can perform a wide range of tasks on behalf of the user, including playing music, booking appointments, and ordering products and services. By incorporating feedback from past interactions, VAs continuously refine their responses and progressively enhance their performance, thus enabling extensive personalization (McLean & Osei-Frimpong, 2019).

Users have the ability to make purchases of products or services exclusively through voice interactions with their VA. Brands like Starbucks, Dunkin' Donuts, and Domino's Pizza have incorporated VAs within their establishments to streamline the process of placing orders for customers. Similarly, Urgent.ly has utilized Amazon's Alexa to empower motorists requiring roadside assistance to initiate calls and make payments. Robust e-commerce platforms such as Amazon and Alibaba have incorporated voice shopping capabilities into their devices (Klaus & Zaichkowsky, 2020). While their popularity and usage for everyday tasks are on the rise, the utilization of VAs for voice shopping is still in its early stages. Only 15% of smart-speaker users in the US regularly make purchases through voice commands (Voicebot.ai, 2021). Although the current usage of VAs for purchasing is limited, there has been a notable surge in the sales of smart speakers during the last years, with a 72% increase in North America alone (Astound Commerce, 2023). The impact of events organized by large companies suggests that voice shopping is a phenomenon yet to be explored (Liao et al., 2019; Vimalkumar et al., 2021). During Alibaba's 24-hour 11.11 shopping festival in China, over one million orders were placed and paid for using Alibaba's Genie voice shopping feature (Li, 2019).

Numerous studies suggest that voice shopping is still in its nascent stage, primarily due to concerns about the technology's ability to ensure user safety, which in turn hinders the development of a trusting relationship between the VA and the user (Bolton et al., 2021; Burns & Igou, 2019). This safety gap introduces an imbalance into the symbolic connection between humans and virtual devices (Ha et al., 2021). However, in comparison to other electronic devices and AI, VAs are known for their unique ability to engage in voice-only conversations and exhibit a somewhat human-like nature. The transmission of voice with human characteristics is a key element of what we refer to as anthropomorphism, and its presence significantly enhances the perception of safety for humans, as demonstrated by the research of Blut et al. (2021) and Liu & Tao (2022). Consequently, a range of studies suggests exploring the influence of anthropomorphism on users' sense of safety when using VAs, particularly in the context of voice shopping.

The literature highlights that users can perceive social emotions when interacting with robots and AI that possess anthropomorphic and human-like attributes, such as VAs capable of speaking and listening (Pelau et al., 2021). However, the concept of anthropomorphism remains fluid, evolving with the state of technology, and definitive studies are yet to provide a precise definition. Notably, when users hear the voice of a computer, they tend to treat it as if it were another person, as evidenced by Nass et al. (1994). Lee & Nass (2003) also showed that their study participants felt compelled to treat the computer as they would any other human being, even while being aware that the computer's voice was synthetic. These studies indicate that interactions between users and VAs, where voice plays a central role, can amplify the effects of anthropomorphism and, potentially, enhance the perceived safety of the device, contributing to a safer voice shopping experience.

Therefore, given these concepts, it becomes imperative to assess the role of anthropomorphism in users' perception of safety associated with this technology, especially as VAs become more sophisticated and acquire more human-like characteristics (Curran & Meuter, 2007; Liu & Tao, 2022). Nevertheless, the impact of anthropomorphism on safety matters remains largely unexplored from an empirical standpoint (Ha et al., 2021; Liu & Tao, 2022). The preceding arguments underscore a critical consideration: certain conditions can either encourage or discourage users from addressing safety-related concerns (Hasan et al., 2021), including their willingness to accept specific trade-offs in exchange for efficient and effective task performance. The extent to which the perception of anthropomorphism

influences perceived safety when using virtual voice assistants to promote voice shopping remains unclear.

This research seeks to expand the existing literature by exploring the concept of anthropomorphism within the realm of VA studies, with a specific focus on safety considerations, alongside the traditional factors of utility and ease. The study delves into two primary dimensions. First, it investigates how anthropomorphic factors, which arise from the human-like qualities of the assistant's voice, impact the perception of safety when using VAs. Second, it aims to quantify the influence of perceived safety on the acceptance of these devices as a viable tool for voice shopping. To precisely define the components of anthropomorphic and human-like characteristics, we have chosen to segment human perception into three distinct aspects: social presence, human-like voice, and friendliness.

The main objective of the study is to investigate the degree to which the assurance of safety when using VAs influences the intention to use them for voice shopping and how this relationship is impacted by the presence of anthropomorphic characteristics in the assistant's voice. In order to validate our proposed theoretical framework and assess the influence of anthropomorphism, we conducted the analysis using three distinct VAs: Alexa, Google Home, and Siri.

2.2. Literature review

2.2.1. Anthropomorphism

Prior studies have consistently emphasized the significance of anthropomorphic characteristics, assessing the extent to which humans interact with non-human technological entities as if they were human, employing human-human interactions as a benchmark (Aggarwal & McGill, 2012; Patrizi et al., 2021). Anthropomorphized cues, such as perceiving a brand as an entity, have been identified as instrumental in understanding anthropomorphism within the realm of consumer research (Aggarwal & McGill, 2012; Patrizi et al., 2021). These insights have created a valuable avenue for researchers exploring VAs (Gao et al., 2007). Additionally, there have been concerted efforts to personalize artificial intelligence (AI) on a more intricate level (Epley, 2018; Festerling & Siraj, 2022), providing a strong foundation for the integration of anthropomorphism in the context of VAs.

Anthropomorphism is the attribution of human-like appearance, character, or behavior to animals, gods, or objects (Cambridge Dictionary, 2023). However, there is currently no unanimous consensus on a singular definition of anthropomorphism, as variations are evident in the literature (Epley, 2018; Festerling & Siraj, 2022).

Early investigations revealed that humans tend to treat computers as if they were human beings when they hear a computer-generated voice (Nass et al., 1994). Subsequent studies expanded on this, finding that participants treated a computer as they would any other human being, even when aware that the voice was synthetic and generated by voice creation software (Lee & Nass, 2003). Furthermore, these studies underscore the potent role of anthropomorphism, especially in interactions where voice takes center stage. With the advancement of AI, VAs have gained the capability to process voice commands and respond in a manner akin to human beings (Russell & Norvig, 2010). Present-day VAs personalize their interactions, allowing for natural conversations that resemble human-to-human exchanges. Unlike other technologies, VAs are progressively evolving toward heightened anthropomorphism, engaging directly with users through voice, as opposed to text and images. This distinctive characteristic of VAs points to an ongoing advancement in which they not only carry out more complex tasks but also interact with other VAs, highlighting the necessity of understanding the safety implications for users and fostering a secure environment in their utilization (Chung et al., 2017; Ha et al., 2021).

It has been observed that a VA's voice exerts a more profound influence on individuals compared to other forms of communication, fostering the relationship between the user and the device (Nass & Yen, 2010). The concept of developing relationships with non-human devices aligns with the theory of parasocial interaction (Horton & Wohl, 1956), which was initially introduced to explain how users could form connections through interactions with various media personas, even when the other participant is not human. This theory has since been extended to describe social relationships between individuals and non-human entities, such as chatbots, robots (Mou & Xu, 2017), and avatars (Fox et al., 2009). The crux of this theory hinges on the perceived humanity of the non-human entity, a fundamental precondition for parasocial interaction, enabling individuals to view the entity as real (Banks & Bowman, 2016).

In the context of VAs, Ha et al. (2021) incorporated the safety aspect into the assistant–user relationship using the Communication privacy management (CPM) theory,

particularly focusing on the high anthropomorphic nature of the interaction in which individuals consider their private information as something they own and control, building barriers around these elements of privacy, which can gradually be deconstructed as the level of safety increases in the relationship with the interlocutor. However, their measurement of anthropomorphism did not include voice-only social presence variables as it was based on a messaging service. While the study established an inverse relationship between risk and anthropomorphism, indicating that greater perceived human presence in the assistant led to reduced risk perception when sharing confidential information, it did not provide a specific conceptualization of anthropomorphism using measurable variables.

To establish a comprehensive framework for defining and quantifying anthropomorphism in the context of VAs, this study identifies, through a literature review, three dimensional constructs that allow defining anthropomorphic presence: Social Presence, Perceived Social Interactions, and Perceived Humanity.

Social Presence: Originating from studies of individuals' subjective perceptions of others in mediated communication, the concept of "social presence" centers on the degree to which a communication medium is perceived as sociable, warm, sensitive, personal, or intimate when used for interaction (Short et al., 1976). While social presence research traditionally focused on human connections mediated by technology, recent studies indicate its applicability to interactions with technological artifacts, such as VAs (McLean & Osei-Frimpong, 2019). Experiments by Reeves & Nass (1997) have demonstrated that individuals unconsciously apply social rules and expectations to computers, with cues related to human characteristics automatically eliciting social responses. This effect has been observed in VA contexts (Chérif & Lemoine, 2019). Furthermore, research has underscored the role of social presence in the adoption of new technologies, especially in building trust in technology. For example, Lu et al. (2016) highlighted the significance of social presence factors in establishing trustworthy and friendly relationships with technology. Verbal communication by robots has been shown to facilitate and enhance perceptions of friendship in interactions with humans (Hildt, 2019).

Social Interactivity: The concept of “social interactivity” revolves around the VA's ability to display appropriate emotions and actions in accordance with social norms. This aspect is particularly crucial when considering the use of a human-like voice in VAs, as it has been extensively studied in contrast to more artificial or robotic voices. The use of a

human-like voice plays a pivotal role in enhancing the VA's trustworthiness and social intelligence. Studies examining the impact of human-like voice on user interactions have predominantly taken place within the domain of human-computer interaction, involving experimental investigations that compare it with more artificial or robotic voices. Notable research conducted by Chattaraman et al. (2019) and Cobos-Guzman et al. (2021) sheds light on the significance of voice in user perceptions. Their findings underscore that the voice of the VA serves as a primary cue for users, influencing their belief that they can interact with the VA in a manner similar to how they would interact with a person. Furthermore, these studies align with broader findings indicating that consumers tend to attribute more anthropomorphic traits to agents equipped with human-like voices. This observation highlights that a human-like voice serves as a precursor to the establishment of social presence, ultimately enhancing the VA's ability to engage users in a more socially interactive and relatable manner. Thus, the use of a human-like voice is a pivotal element in shaping the social interactivity and user perception of virtual assistants.

Perceived Humanity: It is a crucial aspect of user interactions with VAs. It involves the emotions users experience when using a product, often tied to the product's ability to display human-like characteristics such as emotions and gestures. This concept becomes particularly important in the context of friendliness, as VAs are designed to interact in a manner that mimics human behavior (Holzwarth et al., 2006). Friendliness in a VA, influenced by anthropomorphic factors, plays a pivotal role in cultivating friendships and enhancing user interactions. Research studies, such as those conducted by MacInnis & Folkes (2017) and van Doorn et al. (2017), have shown that perceived humanity is a key determinant of VA-user interactions. In situations where consumers are faced with decision-making amid uncertainty, the element of friendliness takes on particular significance. This is especially relevant when individuals must take risks without having complete control over the outcome, as highlighted by Luhmann (2000). As feelings of friendliness increase, consumers tend to perceive a heightened sense of safety in their interactions with VAs, as demonstrated by research conducted by Rousseau et al. (1998). According to Communication Privacy Management (CPM) theory, the presence of signs of friendship and trust between two parties plays a pivotal role in expanding privacy boundaries. This, in turn, fosters a closer relationship, promoting trust and information sharing between the two actors. This concept has been supported by various studies, including those conducted by Kennedy-Lightsey et al. (2012), Mcbride & Bergen (2008) and Petronio et al. (2021). Overall, the perceived humanity

and friendliness of VAs are critical factors in shaping user experiences and the depth of their interactions with these virtual entities.

2.2.2. Safety

The adoption and use of VAs raise unique safety concerns due to the exchange of personal information between users and the assistants. Perceived safety and trust are integral to consumer behavior and become particularly critical in the context of information technology, where sensitive personal data is shared. Given that VAs are a relatively recent, complex, and evolving technology, users may not possess a deep understanding of their inner workings, placing them in a position where they need to trust the technology provider without a comprehensive understanding. This issue has been supported by recent studies that have identified how a lack of trust and understanding can hinder the adoption of new technologies (Hasan et al., 2021). Thus, the adoption of VAs for voice shopping should be examined in light of these insights.

The integration of VAs into everyday life has transformed the way individuals interact with technology. Traditional anonymous interactions with electronic devices no longer apply when using a VA via a smartphone or smart speaker, as these VAs typically possess the user's name and, in many cases, additional personal details due to registration requirements. Moreover, the increasing capabilities of artificial intelligence allow VAs to learn from previous interactions and build comprehensive user profiles. These advancements have given rise to a unique type of safety-based relationship between users and VAs, even when no prior interaction has occurred. This dynamic sets the stage for an interaction quite different from those between individuals (Ha et al., 2021).

Over time, concerns related to safety have escalated, primarily due to data breaches and mismanagement. Various studies have proposed theories to elucidate user behavior in such situations. Trust, information disclosure, and safety concerns have been explored in the context of these new technologies (Bright et al., 2021; Christofides et al., 2009). Cultural factors and mindset variations across domains have been examined through theories such as the construal level theory (Cowan et al., 2021; Krasnova et al., 2012; Liberman & Trope, 1998) and regulatory focus theory (Lwin et al., 2016; Mosteller & Poddar, 2017) in the context of safety concerns and communication.

While these theories are broadly applicable to various technologies, they often lack consideration for a fundamental component characteristic of VAs: their continuous presence in a user's life, whether on a smartphone or a smart speaker. This constant presence establishes a unique personal relationship between users and their VAs. Drawing from Petronio's (2002) Communication Privacy Management (CPM) theory, which addresses user safety concerns in relational contexts, this framework provides valuable insights into privacy issues in user-technology interactions. CPM theory suggests that individuals consider certain information as confidential and, as owners of this private information, seek to maintain control over it. This information encompasses personal thoughts, feelings, experiences, or any other data individuals wish to keep confidential. The theory further posits that individuals set boundaries around their private information to assert control over it (Petronio, 2010).

CPM theory has been applied in various technological contexts, including user disclosure on social media and online forums, where personal information is shared among users and new technology features involving internet connectivity are introduced (Bal et al., 2015; Eastin et al., 2016; Lankton et al., 2017; McNealy & Mullis, 2019; Osatuyi et al., 2018; Waters & Ackerman, 2011). In the context of user-VAs relationships, certain information may be considered confidential, such as credit card numbers, mailing addresses, or purchase histories, while less sensitive requests, like checking the weather or traffic conditions, may be viewed as less privacy-sensitive. The strength of the relationship between the user and the VA influences the user's willingness to share such information. Weaker boundaries may result in the sharing of private information, while stronger boundaries may lead to hesitation in sharing.

Additionally, the Communication Privacy Management theory provides insights into users' intentions to use VAs and their willingness to share information with these entities. When users opt to share information with a VA, they establish what can be termed a "border link." For highly sensitive data, such as purchase history, user addresses, or credit card numbers, users tend to be more willing to share such information with a VA when they perceive a stronger relationship with it (Petronio, 2002). In contrast, when the information's sensitivity is low, as in the case of weather or traffic inquiries, users are generally more open to interacting with any VA, irrespective of the perceived strength of their relationship with the technology (Kim & Kramer, 2015). Thus, this study contends that VAs requesting highly sensitive information may evoke heightened safety-related concerns in users who feel less

closely connected to the VA, in contrast to those who have a stronger sense of familiarity with the technology. Specifically concerning purchases, where the shared information is of a highly sensitive nature, it becomes imperative to fortify the user's connections with the technology. The incorporation of a human-like voice by the assistant emerges as a potential key in reinforcing these connections.

2.2.3. Voice shopping

The retail landscape has witnessed the rise of a novel shopping trend known as voice shopping, facilitated by voice-controlled devices like smart speakers and smartphones equipped with voice interfaces. These voice interfaces can be found in devices with screens, such as smartphones, as well as in screenless devices like Amazon's Echo Dot or Apple's HomePod. It's worth noting that while prior studies have delved into the adoption of VAs, merely owning a VA does not guarantee its utilization for shopping purposes.

Research reveals that many VA owners predominantly employ these devices for non-commercial activities, such as checking the weather or enjoying music (Hu et al., 2022). As a result, theories and empirical insights solely centered on VA adoption might not provide a comprehensive understanding of the emerging and distinctive phenomenon of voice shopping. This highlights the necessity for further exploration into this evolving trend.

Given the growing popularity of voice shopping, numerous theoretical investigations have delved into the factors that propel consumer involvement in this novel retail phenomenon. Simms (2019) shed light on the potential of voice shopping, emphasizing its speed, seamlessness, and the efficient conversion of leads into successful sales when compared to traditional shopping methods. Likewise, Klaus & Zaichkowsky (2022) stressed the pivotal role of convenience and the positive emotions associated with voice assistants as primary drivers of voice shopping. These academic perspectives signify an optimistic outlook for the future of voice shopping, particularly due to rapid technological advancements in this domain.

However, despite these positive perspectives, there remains a lack of clarity regarding the factors that influence the intention to engage in voice shopping. Further research is imperative to unravel the underlying determinants and mechanisms that shape consumer attitudes and behaviors within the context of voice shopping. The existing research gap

underscores the importance of bridging insights from anthropomorphic theories to gain a deeper understanding of user motivations in voice shopping. Addressing this gap can provide valuable insights into the complex relationship between VA users and the driving forces behind their active participation in voice shopping.

Notably, researchers have explored the connection between established theories of social psychology and the effectiveness of specific configurations in virtual assistants. Several theories, including the similarity attraction theory (Byrne & Griffitt, 1969), the emotional contagion theory (Hennig-Thurau et al., 2006), and the dress code theory (Cardon & Okoro, 2009), have been employed to shed light on the relationship between the anthropomorphism of robots and computers and their acceptance by users. Nevertheless, it is the CASA (Computers Are Social Actors) paradigm that underscores that consumers engage with VAs as naturally as they would with other individuals, forming the foundation for contemporary VA research and its acceptance (Nass et al., 1994).

In exploring the impact of anthropomorphism in online shopping, various studies have contributed valuable insights. Jang et al. (2022) focused on investigating the impact of human-robot interaction on consumers' acceptance of humanoid retail service robots, particularly within retail settings. This research aimed to provide insights into the factors influencing consumer acceptance, emphasizing the role played by the interaction between humans and humanoid robots in the realm of retail services. Hu et al. (2022) delved into the relationship between consumers' perceptions of power and their desire for power within the context of voice shopping experiences mediated by AI-powered voice assistants. Their findings unveiled how consumers' perceived power and their desire for power shape choices and preferences during voice shopping interactions, exploring the "fit effect" of these factors.

Additionally, Jang et al. (2022) focused on the concept of smart retailing, with a specific emphasis on anthropomorphism in voice shopping conducted through smart speakers. Their research probed how users attribute human-like characteristics and emotions to AI-driven smart speakers during voice-based shopping interactions and its implications on user behavior and preferences within the context of smart retailing.

In line with these studies, Whang & Im (2021) embarked on an investigation into the influence of human likeness and parasocial relationships on website and voice shoppers' perceptions of product recommendations, guided by the parasocial interaction theory. Their

research scrutinized how users' perceptions of human-like characteristics in AI-driven recommendation systems and their parasocial relationships with these systems affect their responses to product suggestions, across different shopping channels. Rhee & Choi (2020) navigated the territory of personalization and social roles in voice shopping experiences, leveraging the CASA paradigm. Through experiments, they unraveled how product recommendations from a conversational voice agent are influenced by personalization and the assignment of distinct social roles to the agent, shedding light on the significance of these factors in enhancing the effectiveness of voice shopping interactions.

Lastly, Aw et al. (2022) explored the transformative impact of digital voice assistants on customer experiences in voice shopping, anchored in the parasocial theory. Their study highlighted the convenience, personalization, and efficiency that digital voice assistants bring to the voice shopping process, effectively shaping the future of customer experiences.

In conclusion, while existing theories and empirical findings in VA adoption research have made significant contributions, they may not fully encompass the dynamic realm of voice shopping. While convenience and positive emotions have been recognized as key drivers of voice shopping, the precise factors influencing the intention to engage in this shopping method using VAs remain uncertain. Further investigation is required to better grasp the complex relationship between users and the forces that motivate their active participation in voice shopping.

2.3. Conceptual framework and hypotheses development

2.3.1. Anthropomorphic variables

In recent years, VAs have made significant strides in personalizing their interactions with users, allowing for more natural and engaging robot-human conversations. Given that the voice is the defining human characteristic of VAs, the model incorporates three key anthropomorphic variables: Human-Like Voice, social presence, and friendliness. This anthropomorphic aspect of VAs has garnered substantial attention, especially with the proliferation of home-based VAs (Patrizi et al., 2021; Vernuccio et al., 2021).

While some studies have delved into exploring anthropomorphic perceptions of VAs, particularly differentiating human-like voices from robotic voices (Chérif & Lemoine, 2019), the existing literature in this domain appears fragmented. Variables like human-like voices occasionally feature in experimental studies as stimuli without a clear conceptual framework.

Despite the growing interest in anthropomorphism within the VA context, the literature lacks a comprehensive integration of these variables, hindering a holistic understanding of the relationships between users and VAs.

Research on human-like voice in human-computer interaction consistently reveals that users tend to attribute human-like qualities more to computerized agents with human-like voice compared to those using synthetic voices (Schroeder & Epley, 2016). This notion was further supported by Chérif & Lemoine (2019) in VA interactions, emphasizing that human-like voice amplifies users' perception of social presence. Consequently, human-like voice is regarded as a precursor to social presence (Kim et al., 2022; Vilaro et al., 2021).

Research emphasizes the crucial role of social presence in technology adoption and trust-building. Lu et al. (2016) emphasized its importance in fostering trustworthy and amicable relationships with technology, while Hildt (2019) discovered that robots' verbal communication enhances perceptions of friendship in human interactions. Thus, Social Presence is seen as a precursor to friendliness in technology relationships.

Notably, some studies have sought to link the anthropomorphic component with user-perceived safety, primarily within the framework of CPM theory (Chung et al., 2017; Gao et al., 2007; Vilaro et al., 2021; Waters & Ackerman, 2011). However, these studies have yielded contradictory results. Ha et al. (2021) argued that anthropomorphism does not significantly affect the user's perception of safety, while Gao et al. (2007) and Waters & Ackerman (2011) presented arguments to the contrary. Hence, there is a pressing need to establish a clearer foundation within the VA-related literature to ascertain the precise impact of voice anthropomorphism on user privacy boundaries.

In light of the above discussion, we propose the following hypotheses to guide our investigation:

H1. A stronger human perception of the VA's voice is positively related to a heightened sense of social presence.

H2. Greater social presence of the VA positively related with the development of friendliness towards the VA.

H3. A more profound feeling of friendship with the VA is directly linked to an increased sense of safety in the interactions with the VA.

2.3.2. Safety perception

User-perception variables, as highlighted by Tang & Chen (2011), play a pivotal role in shaping user attitudes towards employing voice shopping with VAs, unencumbered by external influences from organizations or companies. Over the years, technology acceptance models have undergone revisions, with certain variables assuming evolving conceptual definitions. These variables serve as instrumental tools for comprehending individuals' proclivity to embrace technology, whether propelled by external or internal factors, such as their inherent disposition to express positivity towards a particular technology (Kaur & Kaur, 2023).

In the realm of VAs, CPM theory, as elucidated in section 2.1, underlines the paramount importance of the safety transmitted by an individual as the foundational key that engenders the requisite confidence to engage in subsequent information exchanges (Petronio, 2002). It's noteworthy that the perception of safety is inherently subjective and varies among individuals, given the variability in their privacy boundaries (Petronio, 2010). For instance, one user might deem it sufficient to entrust their VA, say Alexa, with a username and password to access their events calendar, while another user may extend their trust to the VA to manage their PayPal account for making purchases through the device. Once the privacy boundaries are established and expanded, partly attributed to the perceived humanity of the VA, it results in the fortification of boundary linkages within the assistant-user relationship (Petronio, 2002, 2010). This strengthened union, characterized by an elevated sense of safety, plays a pivotal role in the adoption and endorsement of technology that garners favor due to the human-like attributes of the VA (Chang & Chen, 2014).

Certainly, the research by Arndt (1967) suggests that the perceived safeness, or safety, experienced by consumers can significantly influence their behavior when seeking Word-of-Mouth (WOM) information and making purchase decisions. When individuals feel safe, they are more likely to engage in WOM activities. This indicates that a sense of safety encourages people to seek and share information, participate in discussions about their choices, and express their preferences. Moreover, WOM activity serves not only as a way for individuals to seek information but also as a means for them to distribute information, contributing to the sharing of opinions and recommendations.

The acronym WOM denotes the exchange of information among customers related to a product, service, or business, with the source of information perceived as unaffected by

commercial influence (Litvin et al., 2008). WOM involves a two-way flow of information, where the provider sends information to the seeker (Allsop et al., 2007). Previous studies (Ajzen, 1991; Venkatesh et al., 2003) have established the positive impact of WOM on purchase intention. Given the current low level of technology development, it becomes crucial for those who have already engaged with it to share their experiences.

As previously discussed, safety stands out as a pivotal factor in the adoption of artificial intelligence devices. Hence, a heightened perception of safety among users is essential, as it influences a greater intention to generate positive WOM that can shape the perceptions of potential future users. The sense of safety derived from the humanization of the assistant fosters a feeling of peace of mind in individuals. Research has indicated that when individuals achieve an optimal state in their interaction with technology, the resultant response is an intention to share positive experiences, manifesting as positive WOM (Agag & El-Masry, 2016; Herrando et al., 2019; O'Cass & Carlson, 2010). The sense of safety not only enhances the positive perception among current users but also acts as a catalyst for them to actively propagate their favorable experiences with the technology (Cho, 2019).

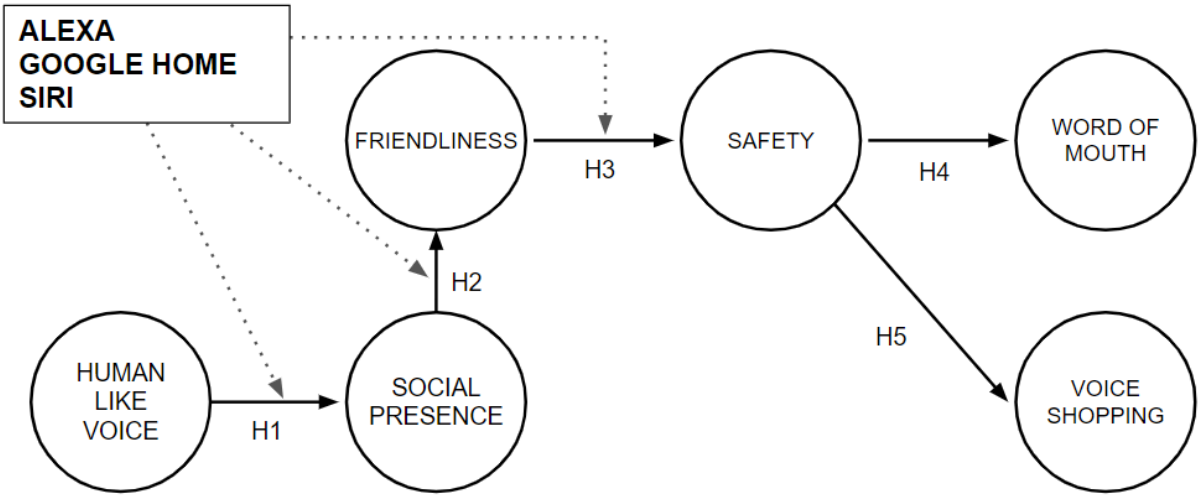
Based on the above, we propose the following hypotheses:

H4. A greater perception of safety in the VA is positively related to the spread of positive WOM.

H5. A greater perception of safety in the VA implies a greater intention to voice shopping.

Figure 2.1 illustrates the proposed model examined here.

Figure 2.1. Proposed model



2.4. Methodology

The data collection process took place in Spain in June 2022 and involved an online survey conducted by a market research company. The participants were selected based on their familiarity with VA technology and the concept of "voice shopping". All of them had had some experience with VAs and knew of their existence. The sample was divided into three equal and representative groups, with each group randomly assigned one of the popular VAs available in the market, namely, Alexa, Google Home, or Siri (Table 2.1). Participants were directed through a web platform specifically designed for this study, where they interacted with audio clips featuring the VAs reading product reviews obtained from the internet. These reviews covered various product categories, including restaurants, books, and cleaning robots.

To ensure exposure to the provided stimuli, all participants began by listening to an introductory audio clip featuring the respective VA discussing their preferred music genre. Subsequently, participants were questioned about the VA's musical preference, and all correctly identified the favorite music genre, indicating their engagement with the stimuli.

Each participant group was then exposed to four audio clips containing current product reviews, narrated in the voice of the VA corresponding to their group. Importantly, respondents were unaware of which assistant's voice they were hearing. These audio clips, lasting approximately 15 to 20 seconds each, were strategically interspersed throughout the survey. To maintain engagement, participants were unable to progress through the survey without listening to these audio clips, as the platform restricted further navigation until the

audio had been played. Additionally, after each audio clip, participants were explicitly asked if they had heard the message clearly, ensuring their attention and comprehension.

To ensure the homogeneity of the groups, the difference in means of the three scenarios (Alexa, Google, and Siri) was analyzed, focusing on variables from the technology acceptance model. Recent research has suggested that the conventional associations between utility and ease of use may have evolved as users become more adept at embracing new technologies (Wu & Wang, 2005). Given the increasing simplicity and ubiquity of voice technologies, particularly in comparison to emerging systems like virtual reality and the metaverse, correlations with user intention to adopt such technologies might have diminished. Consequently, these factors were employed as control variables in the study, with the assumption that participants, considering their sample profiles, should have a consistent level of perception regarding the usefulness and ease of use of VAs.

The subsequent ANOVA analysis confirmed that there were no mean differences among the samples of each of the three scenarios. The reliability of the constructs was analyzed by calculating Cronbach's alpha (Perceived Usefulness: 0.866; Perceived Ease of Use: 0.785). The analysis of ease of use showed no significant group differences (M_{Alexa}: 4.13; M_{Google}: 4.23; M_{Siri}: 4.12; F: 0.409, sig. = 0.664), suggesting that participants perceived the VA technology as equally user-friendly, irrespective of their assigned group. Furthermore, the examination of perceived usefulness did not yield any significant variations between groups (M_{Alexa}: 4.15; M_{Google}: 4.12; M_{Siri}: 4.03; F: 0.446, sig. = 0.640), indicating consistent perceptions of the technology's utility across all participant groups.

These findings instill confidence in the homogeneity of the groups, suggesting that any subsequent differences observed in the study are more likely attributable to the experimental conditions rather than variations in the control variables.

It is important to note that throughout the study, participants remained unaware of which specific VA they were interacting with, thus preserving the study's integrity. The content validity of the survey was ensured by conducting a comprehensive literature review to adapt the variables within the model to the VA context. The survey underwent rigorous scrutiny and validation by multiple experts. All variables in the survey were assessed on a seven-point Likert scale, with responses ranging from 1, indicating strong disagreement, to 7, indicating strong agreement. Prior to the main survey, a pilot study was conducted to confirm respondents' understanding of the survey items.

Table 2.1. Age, gender and VA distribution

Age distribution	Data collected (%)	Gender distribution	Data collected (%)	Virtual assistant distribution	Data collected (%)
18–21 (α)	6.4%	Female	51%	Alexa	33.26%
22–31 (Z)	29%	Male	49%	Google	33.26%
32–41 (Y)	34%	Total	100%	Siri	33.48%
42–50 (X)	31%			Total	100%
Total	100%				

2.5. Results

2.5.1. Testing the measurement model

To examine the proposed hypotheses, we analyzed the collected data via SEM, a sophisticated statistical technique combining multiple regression with confirmatory factor analysis to estimate a series of interrelated dependence relationships simultaneously. In recent years, the number of studies in information systems using SEM to test proposed models has increased (Sarstedt et al., 2022). To ensure the dimensionality, reliability, and validity of the scales, we conducted an exploratory factor and a confirmatory factor analysis using the statistical software SPSS (version 24) and Smart PLS 4. For internal reliability, we measured Cronbach's alpha of all variables exceeding the required minimum value of 0.70, indicating that our variables had satisfactory levels of reliability (Henseler et al., 2016). For convergent validity, item loadings were checked, exceeding the ideal value of 0.70 (Hair, 2009). The composite reliability values of all variables were greater than 0.70, indicating appropriate internal consistency of the variables (Aguirre-Urreta et al., 2013), while the average variance extracted (AVE) showed consistency with values higher or equal to 0.50, indicating that each of the variables captured enough variance over the amount of variance from measurement errors (Fornell & Larcker, 1981). We can conclude that the scales met the requirements of reliability. The results of these analyses are shown in Table 2.2.

Table 2.2. Measurement model results

Variable	Item	Mean	SD	λ	t-value	CA	CR	AVE
HLV	HLV1	3.77	1.77	0.500	174.698***	0.902	0.907	0.910
	HLV2	3.96	1.76	0.548	279.408***			
SPRES	SPRES1	3.75	1.77	0.338	98.571***	0.927	0.930	0.873
	SPRES2	3.50	1.76	0.365	239.022***			
	SPRES3	3.70	1.79	0.367	215.598***			
FRIEND	FRIEND1	4.37	1.49	0.349	153.441***	0.916	0.919	0.856
	FRIEND2	4.22	1.53	0.382	176.890***			
	FRIEND3	4.60	1.53	0.349	120.489***			
SAF	SAF1	4.18	1.69	0.257	66.989***	0.844	0.850	0.679
	SAF2	4.16	1.66	0.266	68.748***			
	SAF3	4.43	1.54	0.354	74.108***			
	SAF4	3.15	1.86	0.342	54.885***			
WOM	WOM1	3.77	1.81	0.363	203.962***	0.923	0.924	0.867
	WOM2	3.65	1.82	0.349	145.754***			
	WOM3	4.07	1.73	0.362	122.128***			
VS	VS1	3.39	1.77	0.389	193.672***	0.879	0.898	0.806
	VS2	3.28	1.86	0.406	201.013***			
	VS3	3.35	1.76	0.313	51.982***			

Notes: CA: Cronbach's alpha; CR: composite reliability; AVE: average variance extracted.
***sig. < 0.01.

HLV: Human-Like Voice; SPRES: Social Presence; FRIEND: Friendliness; SAF: Safety; WOM: Word-of-Mouth; VS: Voice Shopping

We placed a strong emphasis on ensuring discriminant validity, which is vital to confirm that the scales employed in the study effectively represent distinct concepts. To achieve this, we leveraged AVE values (Table 2.2) and conducted a thorough examination of the squared correlation coefficients in the corresponding rows and columns (Table 2.3). This meticulous comparison adhered to the recommended criteria established by Fornell &

Larcker (1981) and Hair (2009). Our efforts in this regard yielded compelling evidence in support of discriminant validity.

Table 2.3. Correlations among factors

	FRIEND	HLV	SAF	SPRES	VS	WOM
FRIEND	0.925					
HLV	0.738	0.954				
SAF	0.663	0.504	0.824			
SPRES	0.766	0.843	0.610	0.934		
VS	0.555	0.453	0.726	0.576	0.898	
WOM	0.693	0.538	0.772	0.665	0.767	0.931

HLV: Human-Like Voice; SPRES: Social Presence; FRIEND: Friendliness; SAF: Safety; WOM: Word-of-Mouth; VS: Voice Shopping

Additionally, we scrutinized the HTMT values, as recommended by Henseler et al. (2016). These values are instrumental in determining the appropriateness of the measurement model, with HTMT values below 0.95 considered suitable primarily for concepts with a high degree of similarity. Our comprehensive assessment, presented in Table 2.4, indicated that all HTMT values remained below the threshold of 0.95, affirming the robustness and validity of our measurement model.

Table 2.4. Heterotrait-monotrait ratio (HTMT)

	FRIEND	HLV	SAF	SPRES	VS	WOM
FRIEND						
HLV	0.81					
SAF	0.730	0.553				
SPRES	0.831	0.919	0.665			
VS	0.608	0.499	0.822	0.629		
WOM	0.752	0.588	0.850	0.720	0.839	

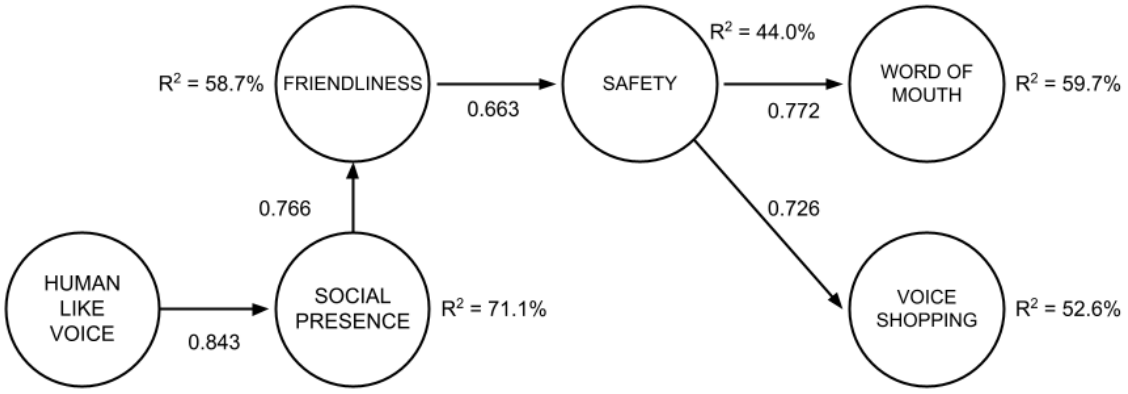
HLV: Human-Like Voice; SPRES: Social Presence; FRIEND: Friendliness; SAF: Safety; WOM: Word-of-Mouth; VS: Voice Shopping

2.5.2. Testing the structural model

We evaluated the structural model's path analysis using a bootstrapping method with 5,000 re-samples. The current structural model explained 52.6% of the voice shopping

variance. The results indicate a satisfactory model as all of the structural model's variance percentages explained were higher than 10% (Falk & Miller, 1992).

Figure 2.2. Estimated model



Note: Significance of all relationships $p < 0.01$.

As anticipated, the contextual variables' hypotheses are all supported (Table 2.5). The study's findings reveal a significant positive relationship between human-like voice and Social Presence (H1), Social Presence and Friendliness (H2), and Friendliness with Safety (H3). These outcomes align with the propositions made by Chérif & Lemoine (2019). Concerning the external variables to technology, all connections are statistically significant. H4, H5, and H6, which assessed the relationship between perceived safety and WOM (H4) and perceived safety with voice shopping, (H5) are supported, corroborating the results of (Chang & Chen, 2014; Hogan et al., 2004).

Table 2.5. Proposed hypotheses

Hypothesis	Dependent variable	Proposed effect	Result
H1	Social presence	Higher with human perception	Supported
H2	Friendliness	Higher with social presence	Supported
H3	Perceived safety	Higher with friendliness	Supported
H4	Word-of-mouth	Higher with perceived safety	Supported
H5	Voice shopping	Higher with perceived safety	Supported

By including the three most popular VAs in the market as moderators of the relationships derived from humanity transmitted through the voice of the assistant (Alexa, Google Home, and Siri), we could analyze the consistency of anthropomorphic relationships while controlling for the type of VA. Table 2.6 provides a detailed comparison through a

bootstrap multigroup analysis of these differences among the VAs. No significant differences were found in the relationships between the three models.

Table 2.6. Bootstrap multigroup analysis results

	Alexa and Google		Google and Siri		Alexa and Siri	
Variables	2-tailed p-value	Diff	2-tailed p-value	Diff	2-tailed p-value	Diff
Human-Like Voice → Social Presence	0.784	No	0.759	No	0.986	No
Social Presence → Friendliness	0.318	No	0.335	No	0.982	No
Friendliness → Safety	0.198	No	0.244	No	0.876	No

2.6. Conclusions and discussion

Our study's primary aim was to develop a theoretical framework that could identify essential anthropomorphic factors influencing the perceived safety in adopting voice shopping through VAs. Specifically, our focus was on the anthropomorphism of VA voices and how these variables align with the CPM theory, offering insights into user behavior when interacting with electronic devices that require the sharing of private and personal information.

Previous research has emphasized the link between anthropomorphism and its influence on safety perception, particularly in technologies handling users' private data. Our research aimed to clarify the role of anthropomorphism in shaping users' safety perceptions during voice shopping interactions, whether through recommendations or transactions. To investigate this, we examined the three predominant VA voices in today's market—Alexa, Google Home, and Siri—to test our hypotheses. The key anthropomorphic variables integrated into our model encompassed the human-like voice, social presence, and friendliness displayed by these VAs. Our study's findings revealed a substantial impact of anthropomorphism on safety perception, not only promoting the acceptance of voice shopping but also its dissemination through word-of-mouth recommendations.

As technology advances, particularly in the realm of voice shopping adoption, the complexity of interactions between users and AI devices intensifies. This increased complexity can potentially harm the user-device relationship by undermining safety perceptions (Gao et al., 2007; Ha et al., 2021). However, when users feel a sense of safety while interacting with VAs, their concerns about personal information exploitation diminish. In line with CPM theory, this strengthened sense of safety weakens privacy boundaries, enabling individuals to share more information and build trust in their interactions with VAs.

Furthermore, the utilization of anthropomorphic variables related to voice also aids in explaining the intention to make favorable recommendations about voice shopping, driven by their ability to assess their satisfaction with the device. Our study's results indicate that the trust conveyed by VAs' voice, which strengthens safety boundaries, facilitates the dissemination of positive information about them. This positive endorsement, coupled with the VAs themselves, results in a heightened intention to recommend voice shopping and potentially engage in voice shopping with them.

As expected, the results supported all hypotheses related to anthropomorphic variables. The results in Table 2.4 and Figure 2.2 show the importance and weight of the anthropomorphic variables related to the assistant's voice, making it possible to analyze the need to continue using human characteristics in these devices. Specifically, the study found that the presence of human-like features in the VA's voice and its mere existence were positive and significantly associated with each other, as well as with users' perceived friendliness towards the VA (H1 and H2, respectively). Additionally, the study found a significant association between users' perceived friendliness and safety towards the VA (H3). These results agree with the three-factor theory of anthropomorphism proposed by Epley (2018), in which they suggested that anthropomorphism can increase the sense of perceived safety by generating feelings of predictability about AI devices. These results also coincide with a study by Pfeuffer et al. (2019), in which the authors showed that the feeling of safety increased in household appliances when they possessed certain human characteristics. The moderating effect of voice type (Alexa, Google and Siri) was analyzed, with no significant differences found. Therefore, we can conclude that the model remains stable for the three analyzed assistants, suggesting that the type of assistant does not influence individuals' perceptions of safety. In all cases, the effect of the perception of humanity in the assistant

remains consistent. Thus, we can conclude that the presence of a voice in the assistant already makes it more human and positively influences the user's perceptions.

Finally, the results show that the perception of security explains the intention to participate in voice shopping and generate active word-of-mouth (H4 and H5). Therefore, a higher perceived security (preceded by an anthropomorphic effect of the voice) is crucial for enabling the adoption of voice shopping and its positive spread. These results align with previous literature (Agag & El-Masry, 2016; O'Cass & Carlson, 2010), suggesting that the intention to express positive word-of-mouth is triggered by a greater sense of security in technology.

2.6.1. Implications for theory

We believe this study makes a valuable contribution to the existing literature on CPM theory, which has been previously applied to various electronic devices and social networks (Lankton et al., 2017; Osatuyi et al., 2018; Waters & Ackerman, 2011). However, studies applying this theory specifically to VAs are limited (Ha et al., 2021). The study can contribute to the literature on CPM theory applied to VAs in a field yet to be exploited—voice shopping. The unique nature of VAs presents a potential risk to many users concerning privacy issues (see Section 2.1). Therefore, we deemed it necessary to explore the factors that could enhance the perceived safety of VAs, aiming to alleviate concerns arising from users' lack of awareness about this technology.

This study contributes significantly to the existing literature on the analysis of anthropomorphism in devices with artificial intelligence. By elucidating the concept of anthropomorphism, this research identifies three variables specifically related to the voice, considering it as the singular human element that VAs possess. The results affirm a positive relationship between human voice perception and social presence, aligning with previous literature support (Kim et al., 2022; Vilaro et al., 2021). The study demonstrates that a heightened perception of the human voice induces a greater sense of social presence, characterized by qualities such as sociability, warmth, and even intimacy, as previously discussed in literature (Short et al., 1976).

These identified characteristics, including sociability and warmth, are deemed essential when establishing and nurturing relationships of trust within the context of the

AI–human connection (Lu et al., 2016). The findings emphasize the importance of these qualities as fundamental pillars in fostering a secure and friendly environment, key elements for a successful AI–human relationship. Therefore, this study not only advances our understanding of anthropomorphism in AI devices but also sheds light on the pivotal role that the human voice plays in shaping perceptions and interactions within the evolving landscape of human-AI relationships.

Unlike studies on CPM theory, the interest that anthropomorphism has aroused in the literature on robots and AI (Li & Sung, 2021; Pelau et al., 2021). Although there has been growing interest in VAs and machine learning in the literature, there is limited knowledge about the nature of interactions and engagement between VAs and consumers. In particular, there is currently no consensus on which anthropomorphic characteristics, particularly those associated with the VA's voice, contribute to generating a stronger sense of friendliness towards the VA.

This approach served two purposes. Firstly, it allowed us to construct a streamlined and focused model, providing a clear and direct path for examining the relationships between other key factors without the potential influence of these two well-established variables. Secondly, by treating Perceived Usefulness and Perceived Ease of Use as control variables, we aimed to control for their impact and verify the homogeneity of our sample. This methodological choice enhances the robustness of our study and ensures that the observed effects are not unduly influenced by these variables.

2.6.2. Implications for practice

The results of this study have practical implications for AI-related research, voice-based software developers, and advertising specialists. Following the recommendations of the previous literature about the perceived risks associated with AI devices, it is essential to be aware of the risks involved in using VAs and how important safety and confidence in the device are considered (Ha et al., 2021). As the relationship with these devices increases, developers and programmers should be aware of how safety can be transmitted to users so that they agree to use them. This study aligns with previous literature, highlighting the fundamental role of VAs in conveying positive messages about their usage. It emphasizes that the humanization of VAs is a crucial aspect in instilling the necessary confidence among users. As such, it is recommended to integrate the use of voice in the communication tools

employed by companies when interacting with their customers. Whether the purpose is to seek information, address inquiries or complaints, or facilitate online purchases, incorporating a humanized voice enhances the overall user experience. By incorporating this human element into communication channels, companies can establish a more trustworthy and engaging connection with their customers, fostering a positive perception of the interactions involving VAs.

The results of this study highlight the importance of perceived security in new technologies. As Petronio (2010) suggested, the removal of barriers between the user and technological devices enables the former to be more inclined to trust personal information to, in this case, the virtual assistant. Consequently, sharing personal information between the user and assistant fosters a greater sense of security in their relationship, thereby facilitating the development of new technologies, such as using virtual assistants for voice shopping.

Given the numerous daily interactions between users and various platforms, providing information about the benefits of using VAs for voice purchases will be crucial for their development. However, the use of VAs for voice purchases is still limited. The main change this technology brings to e-commerce is the ability for consumers to make purchases using voice commands. Therefore, based on the results of this study, we can deduce that the presence of anthropomorphic factors that generate human sensations to the user will allow greater confidence towards the VA, which eventually triggers a more positive attitude about purchasing with these devices. The challenge is to find the right voice that can connect with the user and convey emotions of friendliness, ultimately creating a positive user experience.

2.6.3. Limitations and future research

This research, akin to many other studies, possesses specific limitations. The primary focus of the study centered on general information and purchase searches. To enhance the empirical model's robustness, it would be advantageous to subject it to specific testing for various product or information search categories. One anticipates that the significance of localization could exhibit variance contingent upon the nature of the search task, such as distinguishing between hedonic and utilitarian product searches.

Furthermore, another limitation of the study is its cross-sectional nature. Given the rapid development of technologies and their diverse applications, it is advisable to monitor

the variables examined in this study over time. This longitudinal approach would allow for a more comprehensive understanding of the evolving dynamics and trends in the relationship between the variables, considering the changing landscape of technology and its applications.

As previously highlighted, incorporating social interactivity into a VA acceptance model requires examining the emotions evoked by the perception of the assistant's anthropomorphism. Notably, the type of voice used does not appear to impact the perception of safety. Building on this, a supplementary proposal from the (Marketing Science Institute, 2022) recommends a thorough investigation of emotions using variables with single or double dimensions. Such an exploration could provide insights into the role that emotions play in the process of voice shopping with VAs. As a prospective avenue for future research, we suggest analyzing how the emotions conveyed by each type of voice influence both purchase intention and perceived safety. This analysis can be tailored to consider the type of information required in the transaction, distinguishing between high and low sensitivity contexts.

Rooted in the theoretical framework of Communication Privacy Management, this research endeavor seeks to elucidate the mechanisms by which individuals delineate and regulate boundaries around their private information when interacting with voice-activated shopping assistants. By closely examining how safety factors interact with voice shopping behavior, researchers can uncover valuable insights into how users navigate this complex environment. This inquiry may provide scholarly perspectives on the determinants that mold individuals' propensities to partake in voice shopping, all while taking into account their safety. These findings could carry implications of significant import for both commercial enterprises and the formulation of relevant policies in the continually evolving domain of digital commerce.

2.7. References

Agag, G., & El-Masry, A. A. (2016). Understanding consumer intention to participate in online travel community and effects on consumer intention to purchase travel online and WOM: An integration of innovation diffusion theory and TAM with trust. *Computers in human behavior*, *60*, 97-111.

Aggarwal, P., & McGill, A. L. (2012). When Brands Seem Human, Do Humans Act Like Brands? Automatic Behavioral Priming Effects of Brand Anthropomorphism. *Journal of Consumer Research*, *39*(2), 307-323. https://doi.org/10.1086/662614

Aguirre-Urreta, M. I., Marakas, G. M., & Ellis, M. E. (2013). Measurement of composite reliability in research using partial least squares: Some issues and an alternative approach. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, *44*(4), 11-43. https://doi.org/10.1145/2544415.2544417

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179-211. https://doi.org/10.1016/0749-5978(91)90020-T

Allsop, D. T., Bassett, B. R., & Hoskins, J. A. (2007). Word-of-Mouth Research: Principles and Applications. *Journal of Advertising Research*, *47*(4), 398-411. https://doi.org/10.2501/S0021849907070419

Arndt, J. (1967). Perceived risk, sociometric integration, and word of mouth in the adoption of a new food product. *Risk taking and information handling in consumer behavior*, 289-316.

Astound Commerce. (2023). *Change in smart speaker purchases by region 2021*. Statista. https://www.statista.com/statistics/1276992/change-shopping-via-smart-speaker-worldwide-region/

Aw, E. C.-X., Tan, G. W.-H., Cham, T.-H., Raman, R., & Ooi, K.-B. (2022). Alexa, what's on my shopping list? Transforming customer experience with digital voice assistants.

Technological Forecasting and Social Change, *180*, 121711. https://doi.org/10.1016/j.techfore.2022.121711

Bal, G., Rannenberg, K., & Hong, J. I. (2015). Styx: Privacy risk communication for the Android smartphone platform based on apps' data-access behavior patterns. *Computers & Security*, *53*, 187-202. https://doi.org/10.1016/j.cose.2015.04.004

Banks, J., & Bowman, N. D. (2016). Avatars are (sometimes) people too: Linguistic indicators of parasocial and social ties in player–avatar relationships. *New Media & Society*, *18*(7), 1257-1276. https://doi.org/10.1177/1461444814554898

Bawack, R. E., Wamba, S. F., & Carillo, K. D. A. (2021). Exploring the role of personality, trust, and privacy in customer experience performance during voice shopping: Evidence from SEM and fuzzy set qualitative comparative analysis. *International Journal of Information Management*, *58*, 102309. https://doi.org/10.1016/j.ijinfomgt.2021.102309

Blut, M., Wang, C., Wünderlich, N. V., & Brock, C. (2021). Understanding anthropomorphism in service provision: A meta-analysis of physical robots, chatbots, and other AI. *Journal of the Academy of Marketing Science*, *49*(4), 632-658. https://doi.org/10.1007/s11747-020-00762-y

Bolton, T., Dargahi, T., Belguith, S., Al-Rakhami, M. S., & Sodhro, A. H. (2021). On the Security and Privacy Challenges of Virtual Assistants. *Sensors*, *21*(7), Article 7. https://doi.org/10.3390/s21072312

Bright, L. F., Lim, H. S., & Logan, K. (2021). "Should I Post or Ghost?": Examining how privacy concerns impact social media engagement in US consumers. *Psychology & Marketing*, *38*(10), 1712-1722. https://doi.org/10.1002/mar.21499

Burns, M. B., & Igou, A. (2019). "Alexa, Write an Audit Opinion": Adopting Intelligent Virtual Assistants in Accounting Workplaces. *Journal of Emerging Technologies in*

Accounting, *16*(1), 81-92. https://doi.org/10.2308/jeta-52424

Byrne, D., & Griffitt, W. (1969). Similarity and awareness of similarity of personality characteristics as determinants of attraction. *Journal of Experimental Research in Personality*, *3*(3), 179-186.

Cambridge Dictionary. (2023). Anthropomorphism. En *Cambridge Dictionary*. https://dictionary.cambridge.org/dictionary/english-spanish/anthropomorphism

Cardon, P. W., & Okoro, E. A. (2009). Professional Characteristics Communicated By Formal Versus Casual Workplace Attire. *Business Communication Quarterly*, *72*(3), 355-360. https://doi.org/10.1177/1080569909340682

Chang, C.-H., & Chen, Y.-S. (2014). Managing green brand equity: The perspective of perceived risk theory. *Quality & Quantity*, *48*(3), 1753-1768. https://doi.org/10.1007/s11135-013-9872-y

Charalambous, G., Fletcher, S., & Webb, P. (2016). The Development of a Scale to Evaluate Trust in Industrial Human-robot Collaboration. *International Journal of Social Robotics*, *8*(2), 193-209. https://doi.org/10.1007/s12369-015-0333-8

Chattaraman, V., Kwon, W.-S., Gilbert, J. E., & Ross, K. (2019). Should AI-Based, conversational digital assistants employ social- or task-oriented interaction style? A task-competency and reciprocity perspective for older adults. *Computers in Human Behavior*, *90*, 315-330. https://doi.org/10.1016/j.chb.2018.08.048

Chérif, E., & Lemoine, J.-F. (2019). Anthropomorphic virtual assistants and the reactions of Internet users: An experiment on the assistant's voice. *Recherche et Applications En Marketing (English Edition)*, *34*(1), 28-47. https://doi.org/10.1177/2051570719829432

Cho, E. (2019). Hey Google, Can I Ask You Something in Private? *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1-9.

https://doi.org/10.1145/3290605.3300488

Christofides, E., Muise, A., & Desmarais, S. (2009). Information Disclosure and Control on Facebook: Are They Two Sides of the Same Coin or Two Different Processes? *CyberPsychology & Behavior*, *12*(3), 341-345. https://doi.org/10.1089/cpb.2008.0226

Chung, H., Iorga, M., Voas, J., & Lee, S. (2017). "Alexa, Can I Trust You?" *Computer*, *50*(9), 100-104. Computer. https://doi.org/10.1109/MC.2017.3571053

Cobos-Guzman, S., Nuere, S., De Miguel, L., & König, C. (2021). Design of a Virtual Assistant to Improve Interaction Between the Audience and the Presenter. *International Journal of Interactive Multimedia and Artificial Intelligence*, *7*(2), 232. https://doi.org/10.9781/ijimai.2021.08.017

Cowan, K., Javornik, A., & Jiang, P. (2021). Privacy concerns when using augmented reality face filters? Explaining why and when use avoidance occurs. *Psychology & Marketing*, *38*(10), 1799-1813. https://doi.org/10.1002/mar.21576

Curran, J. M., & Meuter, M. L. (2007). Encouraging Existing Customers to Switch to Self-Service Technologies: Put a Little Fun in their Lives. *Journal of Marketing Theory and Practice*, *15*(4), 283-298. https://doi.org/10.2753/MTP1069-6679150401

Eastin, M. S., Brinson, N. H., Doorey, A., & Wilcox, G. (2016). Living in a big data world: Predicting mobile commerce activity through privacy concerns. *Computers in Human Behavior*, *58*, 214-220. https://doi.org/10.1016/j.chb.2015.12.050

Epley, N. (2018). A Mind like Mine: The Exceptionally Ordinary Underpinnings of Anthropomorphism. *Journal of the Association for Consumer Research*, *3*(4), 591-598. https://doi.org/10.1086/699516

Falk, R. F., & Miller, N. B. (1992). *A primer for soft modeling* (pp. xiv, 103). University of Akron Press.

Festerling, J., & Siraj, I. (2022). Anthropomorphizing Technology: A Conceptual Review of

Anthropomorphism Research and How it Relates to Children's Engagements with Digital Voice Assistants. *Integrative Psychological and Behavioral Science*, *56*(3), 709-738. https://doi.org/10.1007/s12124-021-09668-y

Fornell, C., & Larcker, D. F. (1981). Structural Equation Models with Unobservable Variables and Measurement Error: Algebra and Statistics. *Journal of Marketing Research*, *18*(3), 382-388. https://doi.org/10.2307/3150980

Fowler, G. A. (2018, noviembre 28). Review | I live with Alexa, Google Assistant and Siri. Here's which one you should pick. *Washington Post*. https://www.washingtonpost.com/technology/2018/11/21/i-live-with-alexa-google-assistant-siri-heres-which-you-should-pick/

Fox, J., Bailenson, J., & Binney, J. (2009). Virtual Experiences, Physical Behaviors: The Effect of Presence on Imitation of an Eating Avatar. *Presence: Teleoperators and Virtual Environments*, *18*(4), 294-303. https://doi.org/10.1162/pres.18.4.294

Gao, S., Wang, H., Xu, D., & Wang, Y. (2007). An intelligent agent-assisted decision support system for family financial planning. *Decision Support Systems*, *44*(1), 60-78. https://doi.org/10.1016/j.dss.2007.03.001

Ha, Q.-A., Chen, J. V., Uy, H. U., & Capistrano, E. P. (2021). Exploring the Privacy Concerns in Using Intelligent Virtual Assistants under Perspectives of Information Sensitivity and Anthropomorphism. *International Journal of Human–Computer Interaction*, *37*(6), 512-527. https://doi.org/10.1080/10447318.2020.1834728

Hair, J. (2009). Multivariate Data Analysis. *Faculty and Research Publications*. https://digitalcommons.kennesaw.edu/facpubs/2925

Hasan, R., Shams, R., & Rahman, M. (2021). Consumer trust and perceived risk for voice-controlled artificial intelligence: The case of Siri. *Journal of Business Research*, *131*, 591-597. https://doi.org/10.1016/j.jbusres.2020.12.012

Hennig-Thurau, T., Groth, M., Paul, M., & Gremler, D. D. (2006). Are All Smiles Created Equal? How Emotional Contagion and Emotional Labor Affect Service Relationships. *Journal of Marketing*, *70*(3), 58-73. https://doi.org/10.1509/jmkg.70.3.058

Henseler, J., Hubona, G., & Ray, P. A. (2016). Using PLS path modeling in new technology research: Updated guidelines. *Industrial Management & Data Systems*, *116*(1), 2-20. https://doi.org/10.1108/IMDS-09-2015-0382

Herrando, C., Jiménez-Martínez, J., & Hoyos, M. J. M.-D. (2019). Social commerce users' optimal experience: Stimuli, response and culture. *Journal of Electronic Commerce Research*, *20*(4). https://research.utwente.nl/en/publications/social-commerce-users-optimal-experience-stimuli-response-and-cul

Hildt, E. (2019). Artificial Intelligence: Does Consciousness Matter? *Frontiers in Psychology*, *10*. https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01535

Hogan, J. E., Lemon, K. N., & Libai, B. (2004). Quantifying the Ripple: Word-of-Mouth and Advertising Effectiveness. *Journal of Advertising Research*, *44*(3), 271-280. https://doi.org/10.1017/S0021849904040243

Holzwarth, M., Janiszewski, C., & Neumann, M. M. (2006). The Influence of Avatars on Online Consumer Shopping Behavior. *Journal of Marketing*, *70*(4), 19-36. https://doi.org/10.1509/jmkg.70.4.019

Horton, D., & Wohl, R. (1956). Mass Communication and Para-Social Interaction. *Psychiatry*, *19*(3), 215-229. https://doi.org/10.1080/00332747.1956.11023049

Hu, P., Lu, Y., & Wang, B. (2022). Experiencing power over AI: The fit effect of perceived power and desire for power on consumers' choice for voice shopping. *Computers in Human Behavior*, *128*, 107091. https://doi.org/10.1016/j.chb.2021.107091

Jang, Y.-T. J., Liu, A. Y., & Ke, W.-Y. (2022). Exploring smart retailing: Anthropomorphism

in voice shopping of smart speaker. *Information Technology & People*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/ITP-07-2021-0536

Kaur, D., & Kaur, R. (2023). Does electronic word-of-mouth influence e-recruitment adoption? A mediation analysis using the PLS-SEM approach. *Management Research Review*, *46*(2), 223-244. https://doi.org/10.1108/MRR-04-2021-0322

Kennedy-Lightsey, C. D., Martin, M. M., Thompson, M., Himes, K. L., & Clingerman, B. Z. (2012). Communication Privacy Management Theory: Exploring Coordination and Ownership Between Friends. *Communication Quarterly*, *60*(5), 665-680. https://doi.org/10.1080/01463373.2012.725004

Kim, H. C., & Kramer, T. (2015). Do Materialists Prefer the "Brand-as-Servant"? The Interactive Effect of Anthropomorphized Brand Roles and Materialism on Consumer Responses. *Journal of Consumer Research*, *42*(2), 284-299. https://doi.org/10.1093/jcr/ucv015

Kim, J., Merrill Jr., K., Xu, K., & Kelly, S. (2022). Perceived credibility of an AI instructor in online education: The role of social presence and voice features. *Computers in Human Behavior*, *136*, 107383. https://doi.org/10.1016/j.chb.2022.107383

Klaus, P., & Zaichkowsky, J. (2020). AI voice bots: A services marketing research agenda. *Journal of Services Marketing*, *34*(3), 389-398. https://doi.org/10.1108/JSM-01-2019-0043

Klaus, P., & Zaichkowsky, J. L. (2022). The convenience of shopping via voice AI: Introducing AIDM. *Journal of Retailing and Consumer Services*, *65*, 102490. https://doi.org/10.1016/j.jretconser.2021.102490

Krasnova, H., Veltri, N. F., & Günther, O. (2012). Die Rolle der Kultur in der Selbstoffenbarung und Privatsphäre in sozialen Onlinenetzwerken. *WIRTSCHAFTSINFORMATIK*, *54*(3), 123-133.

https://doi.org/10.1007/s11576-012-0323-5

Lankton, N. K., McKnight, D. H., & Tripp, J. F. (2017). Facebook privacy management strategies: A cluster analysis of user privacy behaviors. *Computers in Human Behavior*, *76*, 149-163. https://doi.org/10.1016/j.chb.2017.07.015

Lee, K. M., & Nass, C. (2003). Designing Social Presence of Social Actors in Human Computer Interaction. *NEW HORIZONS*, *5*, 8.

Li, C. (2019, noviembre 12). *Millions in China Tapped «Voice Shopping» During 11.11*. Alizila. https://www.alizila.com/millions-in-china-tapped-voice-shopping-during-11-11-tmall-genie/

Li, X., & Sung, Y. (2021). Anthropomorphism brings us closer: The mediating role of psychological distance in User–AI assistant interactions. *Computers in Human Behavior*, *118*, 106680. https://doi.org/10.1016/j.chb.2021.106680

Liao, Y., Vitak, J., Kumar, P., Zimmer, M., & Kritikos, K. (2019). Understanding the Role of Privacy and Trust in Intelligent Personal Assistant Adoption. En N. G. Taylor, C. Christian-Lamb, M. H. Martin, & B. Nardi (Eds.), *Information in Contemporary Society* (pp. 102-113). Springer International Publishing. https://doi.org/10.1007/978-3-030-15742-5_9

Liberman, N., & Trope, Y. (1998). The role of feasibility and desirability considerations in near and distant future decisions: A test of temporal construal theory. *Journal of Personality and Social Psychology*, *75*(1), 5-18. https://doi.org/10.1037/0022-3514.75.1.5

Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, *29*(3), 458-468. https://doi.org/10.1016/j.tourman.2007.05.011

Liu, K., & Tao, D. (2022). The roles of trust, personalization, loss of privacy, and anthropomorphism in public acceptance of smart healthcare services. *Computers in Human Behavior*, *127*, 107026. https://doi.org/10.1016/j.chb.2021.107026

Lu, B., Fan, W., & Zhou, M. (2016). Social presence, trust, and social commerce purchase intention: An empirical research. *Computers in Human Behavior*, *56*, 225-237. https://doi.org/10.1016/j.chb.2015.11.057

Luhmann, N. (2000). Familiarity, Confidence, Trust: Problems and Alternatives. *Trust : Making and peaking Cooperative Relations*. https://cir.nii.ac.jp/crid/1573950400009493504

Lwin, M. O., Wirtz, J., & Stanaland, A. J. S. (2016). The privacy dyad: Antecedents of promotion- and prevention-focused online privacy behaviors and the mediating role of trust and privacy concern. *Internet Research*, *26*(4), 919-941. https://doi.org/10.1108/IntR-05-2014-0134

MacInnis, D. J., & Folkes, V. S. (2017). Humanizing brands: When brands seem to be like me, part of me, and in a relationship with me. *Journal of Consumer Psychology*, *27*(3), 355-374. https://doi.org/10.1016/j.jcps.2016.12.003

Marketing Science Institute. (2022). *Research Priorities 2022-2024*. MSI - Marketing Science Institute. https://www.msi.org/article/msi-announces-2022-24-research-priorities/

Mcbride, C., & Bergen, K. (2008). Becoming a Reluctant Confidant: Communication Privacy Management in Close Friendships. *Texas Journal of Speech Communication*, *33*, 50-61.

McLean, G., & Osei-Frimpong, K. (2019). Hey Alexa … examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in Human Behavior*, *99*, 28-37. https://doi.org/10.1016/j.chb.2019.05.009

McNealy, J., & Mullis, M. D. (2019). Tea and turbulence: Communication privacy management theory and online celebrity gossip forums. *Computers in Human Behavior*, *92*, 110-118. https://doi.org/10.1016/j.chb.2018.10.029

Moriuchi, E. (2019). Okay, Google!: An empirical study on voice assistants on consumer engagement and loyalty. *Psychology & Marketing*, *36*(5), 489-501. https://doi.org/10.1002/mar.21192

Mosteller, J., & Poddar, A. (2017). To Share and Protect: Using Regulatory Focus Theory to Examine the Privacy Paradox of Consumers' Social Media Engagement and Online Privacy Protection Behaviors. *Journal of Interactive Marketing*, *39*(1), 27-38. https://doi.org/10.1016/j.intmar.2017.02.003

Mou, Y., & Xu, K. (2017). The media inequality: Comparing the initial human-human and human-AI social interactions. *Computers in Human Behavior*, *72*, 432-440. https://doi.org/10.1016/j.chb.2017.02.067

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 72-78. https://doi.org/10.1145/191666.191703

Nass, C., & Yen, C. (2010). *The Man Who Lied to His Laptop: What We Can Learn About Ourselves from Our Machines*. Penguin.

O'Cass, A., & Carlson, J. (2010). Examining the effects of website-induced flow in professional sporting team websites. *Internet research*, *20*(2), 115-134.

Osatuyi, B., Passerini, K., Ravarini, A., & Grandhi, S. A. (2018). "Fool me once, shame on you… then, I learn." An examination of information disclosure in social networking sites. *Computers in Human Behavior*, *83*, 73-86. https://doi.org/10.1016/j.chb.2018.01.018

Patrizi, M., Vernuccio, M., & Pastore, A. (2021). "Hey, voice assistant!" How do users

perceive you? An exploratory study. *Sinergie Italian Journal of Management*, *39*(1), Article 1. https://doi.org/10.7433/s114.2021.10

Pelau, C., Dabija, D.-C., & Ene, I. (2021). What makes an AI device human-like? The role of interaction quality, empathy and perceived psychological anthropomorphic characteristics in the acceptance of artificial intelligence in the service industry. *Computers in Human Behavior*, *122*, 106855. https://doi.org/10.1016/j.chb.2021.106855

Petronio, S. (2002). *Boundaries of Privacy: Dialectics of Disclosure*. SUNY Press.

Petronio, S. (2010). Communication Privacy Management Theory: What Do We Know About Family Privacy Regulation? *Journal of Family Theory & Review*, *2*(3), 175-196. https://doi.org/10.1111/j.1756-2589.2010.00052.x

Petronio, S., Child, J. T., & Hall, R. D. (2021). Communication Privacy Management Theory: Significance for Interpersonal Communication. En *Engaging Theories in Interpersonal Communication* (3.ª ed.). Routledge.

Pfeuffer, N., Benlian, A., Gimpel, H., & Hinz, O. (2019). Anthropomorphic Information Systems. *Business & Information Systems Engineering*, *61*(4), 523-533. https://doi.org/10.1007/s12599-019-00599-y

Reeves, B., & Nass, C. (1997). The Media Equation: How People Treat Computers, Television, ? New Media Like Real People ? Places. *Computers and Mathematics with Applications*, *5*(33), 128.

Rhee, C. E., & Choi, J. (2020). Effects of personalization and social role in voice shopping: An experimental study on product recommendation by a conversational voice agent. *Computers in Human Behavior*, *109*, 106359. https://doi.org/10.1016/j.chb.2020.106359

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not So Different After All:

A Cross-Discipline View Of Trust. *Academy of Management Review*, *23*(3), 393-404. https://doi.org/10.5465/amr.1998.926617

Russell, S., & Norvig, P. (2010). *Intelligence artificielle: Avec plus de 500 exercices*. Pearson Education France.

Sarkar, S., Chauhan, S., & Khare, A. (2020). A meta-analysis of antecedents and consequences of trust in mobile commerce. *International Journal of Information Management*, *50*, 286-301. https://doi.org/10.1016/j.ijinfomgt.2019.08.008

Sarstedt, M., Hair, J. F., Pick, M., Liengaard, B. D., Radomir, L., & Ringle, C. M. (2022). Progress in partial least squares structural equation modeling use in marketing research in the last decade. *Psychology & Marketing*, *39*(5), 1035-1064. https://doi.org/10.1002/mar.21640

Schroeder, J., & Epley, N. (2016). Mistaking minds and machines: How speech affects dehumanization and anthropomorphism. *Journal of Experimental Psychology: General*, *145*(11), 1427-1437. https://doi.org/10.1037/xge0000214

Short, R. V., Short, R. V., & Baird, D. T. (1976). Definition of the problem—The evolution of human reproduction. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *195*(1118), 3-24. https://doi.org/10.1098/rspb.1976.0095

Simms, K. (2019). How voice assistants could change the way we shop. *Harvard Business Review*, *may*, 2-7.

Tang, D., & Chen, L. (2011). A review of the evolution of research on information Technology Acceptance Model. *2011 International Conference on Business Management and Electronic Information*, *2*, 588-591. https://doi.org/10.1109/ICBMEI.2011.5917980

van Doorn, J., Mende, M., Noble, S. M., Hulland, J., Ostrom, A. L., Grewal, D., & Petersen, J. A. (2017). Domo Arigato Mr. Roboto: Emergence of Automated Social Presence in

Organizational Frontlines and Customers' Service Experiences. *Journal of Service Research*, *20*(1), 43-58. https://doi.org/10.1177/1094670516679272

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *Vol. 27*(No. 3), 425-478.

Vernuccio, M., Patrizi, M., & Pastore, A. (2021). Delving into brand anthropomorphisation strategies in the experiential context of name‐brand voice assistants. *Journal of Consumer Behaviour*, cb.1984. https://doi.org/10.1002/cb.1984

Vilaro, M. J., Wilson-Howard, D. S., Zalake, M. S., Tavassoli, F., Lok, B. C., Modave, F. P., George, T. J., Odedina, F., Carek, P. J., & Krieger, J. L. (2021). Key changes to improve social presence of a virtual health assistant promoting colorectal cancer screening informed by a technology acceptance model. *BMC Medical Informatics and Decision Making*, *21*(1), 196. https://doi.org/10.1186/s12911-021-01549-z

Vimalkumar, M., Sharma, S. K., Singh, J. B., & Dwivedi, Y. K. (2021). 'Okay google, what about my privacy?': User's privacy perceptions and acceptance of voice based digital assistants. *Computers in Human Behavior*, *120*, 106763. https://doi.org/10.1016/j.chb.2021.106763

Voicebot.ai. (2021). *Voice Shopping Consumer Adoption Report 2021*. https://voicebot.ai/2021/12/24/voice-shopping-rises-to-45-million-u-s-adults-in-2021/

Waters, S., & Ackerman, J. (2011). Exploring Privacy Management on Facebook: Motivations and Perceived Consequences of Voluntary Disclosure. *Journal of Computer-Mediated Communication*, *17*(1), 101-115. https://doi.org/10.1111/j.1083-6101.2011.01559.x

Whang, C., & Im, H. (2021). «I Like Your Suggestion!» the role of humanlikeness and parasocial relationship on the website versus voice shopper's perception of

recommendations. *Psychology & Marketing*, *38*(4), 581-595. https://doi.org/10.1002/mar.21437

Wu, J.-H., & Wang, S.-C. (2005). What drives mobile commerce?: An empirical evaluation of the revised technology acceptance model. *Information & Management*, *42*(5), 719-729. https://doi.org/10.1016/j.im.2004.07.001

Zhong, B., & Yang, F. (2018). How We Watch TV Tomorrow?: Viewers' Perception Towards Interactivity on Smart TV. *International Journal of Asian Business and Information Management (IJABIM)*, *9*(4), 48-63. https://doi.org/10.4018/IJABIM.2018100104

2.8. Appendix

Table 2.7. Scales

Item	Scales
Human Like Voice - Adapted from Patrizi et al. (2021)	
HLV1	The assistant's voice sounds human
HLV2	The quality of his voice seems human
Social Presence - Adapted from McLean & Osei-Frimpong (2019) and Patrizi et al. (2021)	
SPRES1	I would interact with this VA as if I were human
SPRES2	When I hear the voice of this VA I feel like I am listening to a person
SPRES3	I feel a human presence in their voice
Friendliness - Adapted from Charalambous et al. (2016)	
FRIEND1	The assistant's voice makes me feel comfortable
FRIEND2	The assistant's voice would make me trust them
FRIEND3	I would feel comfortable interacting with this VA
Safety - Adapted from McLean & Osei-Frimpong (2019)	
SAF1	I feel safe using VA
SAF2	I feel that VAs are a safe technology
SAF3	I would feel safe using this VA
SAF4	I would trust personal information to this VA
Word of mouth - Adapted from Moriuchi (2019)	
WOM1	I would recommend friends and family to buy with this VA
WOM2	I would encourage my close circle to voice shopping by using a VA with this voice
WOM3	I would speak well to my close circle of a VA with this voice
Perceived Usefulness - Adapted from Moriuchi (2019)	
PU1	Using a VA would make it easier to find information about my purchases on the internet
PU2	Using a VA would save me time when looking for information about a product or service

PU3	I believe that the use of VAs is useful to buy online
Perceived Ease of Use - Adapted from Sarkar et al. (2020)	
PEOU1	VAs are easy to use
PEOU2	VA technology is intuitive
Voice Shopping - Adapted from Bawack et al. (2021)	
VSI1	In the future I will continue to use VA in my daily life (eg.: music, weather, shopping…)
VSI2	In the future I will continue to use VA to search for purchasing information
VSI3	In the future I will use VA to shop online

STUDY 3

Trusting Virtual Assistants: The Power of Emotional Design

3.1. Introduction

Voice assistants (VAs) have quickly reshaped how users engage with digital technology, establishing themselves as integral to daily life. By the end of 2024, it is estimated that more than 8.4 billion digital voice assistants will be in use globally—outnumbering the world's population (Statista Staff, 2023). Organizations are increasingly leveraging these technologies not only to reduce costs but also to enhance user engagement. The timing of adopting innovative capabilities, such as VAs, plays a critical role in influencing organizational performance, particularly in competitive sectors like e-commerce (Fuller et al., 2022). The emotional expressiveness of synthetic voices could be a critical factor in strengthening brand equity and customer loyalty (Huh et al., 2023). This reflects their role in managing tasks, providing information, and facilitating interactions. VAs utilize natural language processing (NLP) and machine learning to understand user inputs, improving over time by adapting to nuances such as accents and emotional cues (Coker & Thakur, 2024). In business, VAs streamline customer service, schedule management, and data retrieval, enhancing efficiency and cutting costs (Pfister & Lehmann, 2021). Their ability to evoke human-like emotions further boosts satisfaction and loyalty by creating a sense of warmth and trust (Cheshin et al., 2018; Gelbrich et al., 2021).

Emotional Response (ER), involving valence (positivity or negativity) and arousal (intensity), is key to engaging interactions (Russell & Barrett, 1999). Basic Emotion Theory identifies emotions like happiness and anger, each with distinct valence and arousal levels (Ekman, 1992). A VA's simulated human voice directly shapes ER, influencing trust and user perception (Cuciniello et al., 2023). Human-like voices enhance interaction quality and trust in business contexts (Calahorra-Candao & Martín-de Hoyos, 2024). Research shows that a more human-sounding VA is seen as more empathetic and trustworthy, encouraging user reliance in settings like customer service and online shopping (Huh et al., 2023; Schroeder & Epley, 2015).

Humanizing synthetic voices is crucial for making VAs more relatable. By adding emotional expressiveness, VAs can improve user attitudes and behaviors, making interactions feel more human (Poushneh, 2021). Users are more inclined to trust systems that exhibit human traits, especially in trust-dependent fields like customer service and e-commerce (Cowan et al., 2015; Schroeder & Epley, 2015). Designing voices that convey emotions

boosts satisfaction, trust, and loyalty, enhancing VAs' effectiveness (Shin et al., 2021).

Users' emotional responses to VAs depend on tone and warmth conveyed by the voice (Pepino et al., 2020). A warm, calm voice with positive valence and moderate arousal fosters trust and satisfaction (Schroeder & Epley, 2015). In contrast, harsh or intense voices with negative valence and high arousal can cause discomfort, reducing trust and engagement (Russell, 1980). Research shows users are more likely to trust VAs that evoke emotional responses, making them seem more empathetic and relatable (Schroeder & Epley, 2015). However, synthetic voices often lack consistency in emotional cues, limiting connections with users (Lucia-Palacios & Pérez-López, 2021). Familiar voice qualities improve user enjoyment and trust, particularly in customer service settings where trust is crucial (Hall et al., 2024). Enhancing ER can significantly boost trust and interaction quality (Chin & Yi, 2022).

While VAs perform tasks efficiently, they often lack the emotional expressiveness needed for human-like, trustworthy interactions (Huang & Sénécal, 2023). Mainly designed to recognize commands rather than evoke complex emotions, they fall short in contexts requiring trust, like customer service (Coker & Thakur, 2024). Improving VAs' ability to evoke suitable emotional responses could enhance trust and emotional intelligence.

While previous research has focused on the appearance and conversational style of VAs, the role of their voice in eliciting emotional responses has been comparatively overlooked. Studies on VA anthropomorphism often emphasize physical design, demonstrating that human-like appearances enhance user trust and engagement (Gong & Nass, 2007; Krämer et al., 2015). Similarly, research on conversational style highlights how politeness and personalization positively impact trust (Nass & Moon, 2000; Reeves & Nass, 1997). Emotional tone in dialogues has also been explored, but these studies primarily focus on text-based or broader conversational attributes rather than the expressiveness of synthetic voices (Brave & Nass, 2007). Although voice-related research has examined general vocal traits like pitch and speech rate and their influence on trustworthiness, the specific role of emotional expressiveness in synthetic voices remains underexplored. This study addresses this gap by analyzing how VAs' ability to convey emotions through voice impacts trust and human-likeness.

Emotional marketing has shown that sensory stimuli, such as distinctive scents,

effectively strengthen the bond between consumers and brands, enhancing trust and memorability (Cheshin et al., 2018; Gelbrich et al., 2021). Similarly, synthetic voices with carefully designed emotional attributes have the potential to evoke trust and reinforce brand recognition. This study examines how these emotional dimensions influence the perception of virtual assistants and their effectiveness in building human-like connections with users (Fakhimi et al., 2023; Huh et al., 2023).

This study addresses a gap in emotional expressiveness in VAs, focusing on their ability to trigger basic emotions in users. By examining how ER, influenced by valence and arousal, shapes perceptions of a VA's humanity, this research offers insights for creating emotionally intelligent VAs. Moving beyond functionality, it redefines VAs as systems capable of fostering emotional connections, enhancing perceptions of humanity and trust.

The main objective is to analyze how synthetic voices convey emotions that shape anthropomorphism and user trust, identifying which voices foster the most human-like interactions. By exploring both emotional dimensions and responses, this research deepens understanding of how to integrate emotional intelligence into VAs (Ramesh & Chawla, 2022). The study examines VAs like Siri, Google Assistant, and Alexa, using an experimental approach with 900 participants exposed to product reviews (positive and negative), providing comprehensive insights into emotional responses and trust.

This study significantly contributes to human-computer interaction, AI development, and VA applications by examining how ER influences user trust. While much literature discusses enhancing VA functionality, few studies focus on how VAs generate emotional cues that affect user perceptions (Fakhimi et al., 2023; McLean & Osei-Frimpong, 2019). By focusing on the human perspective, this research fills a critical gap, revealing how VAs' emotional expressiveness impacts user trust and perception, advancing HCI research.

For AI development, this study underscores the need for emotionally intelligent systems that elicit user emotions, with implications for NLP and speech synthesis technologies. From a practical perspective, it provides valuable insights for developers and designers, helping them align VAs with user expectations in sensitive areas like customer service, healthcare, and education.

3.2. Theoretical framework

3.2.1. Emotional response to basic emotions

Basic emotions are generally categorized as positive or negative, each exerting distinct effects on user perception. Positive emotions—such as happiness, calmness, and empathy—tend to foster greater satisfaction and trust in human-computer interaction, while negative emotions—like anger, frustration, and sadness—can lead to discomfort, reducing both trust and user satisfaction (Russell & Barrett, 1999). Emotional appeals have also been shown to heighten attention and emotional involvement, highlighting their persuasive power in digital environments (Sandoval & García-Madariaga, 2024). Recognizing this emotional duality is essential for understanding how users respond to different stimuli, and underscores the importance of carefully balancing positive and negative expressions in synthetic voice design (Herrando et al., 2022; Qu et al., 2014). Marketing research also highlights emotions' impact on consumer behavior, viewing them as central to decision-making and trust (Russell & Barrett, 1999). Studies suggest that positive emotional responses are crucial for building trust in digital interactions, with VAs that convey positive emotions perceived as supportive, approachable, and more human-like (McLean & Osei-Frimpong, 2019).

Negative emotions, while often less desirable, add realism to VA interactions by mimicking authentic human behavior. When a VA expresses emotions like disappointment, it can enhance perceived humanity, as users associate a full range of emotions with real interactions (Mostafa et al., 2024). However, the intensity of these expressions should be carefully managed, as overly negative responses can reduce trust and increase anxiety (Zhang et al., 2021). This balance is especially crucial in professional settings where reliability and professionalism are key.

In business contexts, where VAs are used in customer service, product recommendations, and interactions, emotional expressiveness shapes user experience. However, extremes in expression, like overly enthusiastic or desperate tones, may negatively impact trust and professionalism. Overly positive or negative emotional tones can disrupt user trust, leading to discomfort or perceptions of inauthenticity (Fakhimi et al., 2023). Such extremes are especially problematic in professional environments, where qualities like professionalism, clarity, and neutrality are valued.

Expressions of desperation or exaggerated cues can elevate user anxiety, reducing a

VA's effectiveness in professional roles. Emotional responses that deviate from norms may harm customer satisfaction and erode trust (Mostafa et al., 2024). In contrast, balanced emotions—such as warmth, empathy, or calmness without extremes—are more likely to foster trust and loyalty (Huang & Sénécal, 2023). Flow theory, often applied in HCI, suggests users achieve satisfaction when emotional states match a balanced level of challenge and skill (Csikszentmihalyi, 1997). For VAs, this implies that emotional expressiveness should enhance satisfaction without overwhelming users, especially in business environments where emotional regulation is essential (Huang & Sénécal, 2023).

The combination of basic positive and negative emotions creates an ER that influences user behavior. ER is the emotional reaction to stimuli—like sounds or colors—based on perception and interpretation, defined by valence (positivity or negativity) and arousal (intensity) (Russell & Barrett, 1999). For example, a pleasant sound might evoke positive valence and moderate arousal, fostering comfort and trust, while a sudden loud noise might induce negative valence and high arousal, leading to discomfort or fear.

Negative emotional expressions, in turn, may trigger adverse responses, reducing satisfaction, trust, and the perception of a VA's humanity (Qu et al., 2014). This suggests that VAs conveying negative emotions can evoke negative responses, potentially undermining interaction quality and diminishing the VA's human-like appeal.

Therefore, in light of the literature review, we propose the following hypotheses:

H1a. The voice produced by virtual assistants has the ability to convey basic positive emotions that positively impact the user's emotional response.

H1b. The voice produced by virtual assistants has the ability to convey basic negative emotions that negatively impact the user's emotional response.

3.2.2. Human-like voice

Emotional reactions to a virtual assistant’s voice play a key role in shaping its perceived humanity. According to Emotional Response Theory, users are more likely to attribute human-like qualities to a system when it evokes emotion—particularly when the voice conveys warmth, calmness, or enthusiasm, which are closely associated with human social interaction (Fakhimi et al., 2023; Russell & Barrett, 1999). This emotional connection

also underpins trust. Neuroscientific evidence shows that exposure to digital trust cues activates brain regions involved in risk assessment and decision-making, suggesting that trust is not purely cognitive but also rooted in emotional and subconscious processing (Casado-Aranda et al., 2019). Together, these findings support the idea that emotionally expressive voice assistants can foster both perceived humanity and trust on a deeper, non-rational level.

Research on anthropomorphism emphasizes emotional expressiveness as crucial for shaping perceptions of humanity in digital agents. Other studies have extended this perspective, showing that emotionally engaging intelligent assistants foster deeper connections, bridging functional utility and relational bonding (Song et al., 2022), and others have indicated that virtual assistants with voices are seen as more human and relatable (Calahorra-Candao & Martín-de Hoyos, 2024). Emotional cues in these voices suggest an understanding of human emotions, bridging the gap between artificial systems and human presence (Schroeder & Epley, 2015).

Social presence theory supports the idea that a virtual assistant's perceived humanity is enhanced when it triggers emotional responses. When users experience emotional reactions—such as arousal and valence—they are more likely to unconsciously attribute social and human-like qualities to the assistant, making the interaction feel more personal and authentic (Wu et al., 2023). This psychological mechanism is further reinforced by neurophysiological evidence showing that unconscious responses to sensory stimuli can significantly influence consumer decision-making, underscoring the importance of emotionally expressive voice interfaces (García-Madariaga et al., 2020).

H2. The activation of the user's arousal and valence as a consequence of the emotions perceived in the VA's voice positively increases the perception of the VA's humanity.

3.2.3. Trust in VAs

Trust is essential in VA interactions, particularly as these systems integrate into daily tasks and decision-making. The emotional tone of synthetic voices greatly influences user perception and interaction, creating reliability and fostering loyalty (Fakhimi et al., 2023). Voice integration in VAs enhances their perception as more human, which increases user acceptance (Calahorra-Candao & Martín-de Hoyos, 2024). Human-like qualities bridge

emotional responses with user trust, as conversational AI can substitute for trust, with the onboarding experience shaping perceptions and decisions (Hildebrand & Bergner, 2021). Users perceive a VA's voice as more relatable and trustworthy when it exhibits warmth, naturalness, and the ability to express emotions (Pagani et al., 2019). Voice interactions in omnichannel strategies have been shown to boost brand trust (Malodia et al., 2023). AI-driven voice assistants enhance customer service and satisfaction, making VAs feel more authentic and engaging (Schroeder & Epley, 2015)..

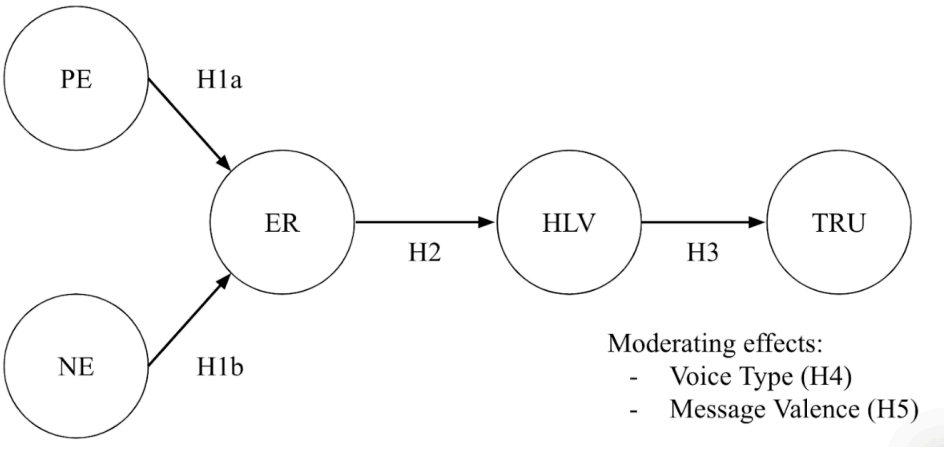
Empathic responses from VAs enhance the user experience, making interactions feel human-like. When VAs are emotionally expressive, users are more inclined to trust their suggestions, as perceived emotional intelligence builds trust (Paradeda et al., 2024). Poushneh (2021) also found that human-like personalities in VAs positively affect consumer attitudes and trust. The Computational Model of Trust by Lee & See (2004) suggests that trust in HCI involves performance, process, and purpose. In VAs, performance trust refers to task effectiveness (Zhang et al., 2021), while process trust includes tone and expressiveness. Purpose trust relates to alignment with user needs, which empathic VAs reinforce, showing they act in the user's best interest.

Studies indicate that natural and emotionally rich synthetic voices strengthen user trust, especially in long-term interactions (Schroeder & Epley, 2015). Huh et al. (2023) found that AI-powered VAs foster brand loyalty through trust and satisfaction. However, the literature lacks an in-depth focus on how ER in synthetic voice shapes human-likeness, crucial for VA trust. Coker & Thakur (2024) noted that many VAs lack emotionally engaging voices, resulting in transactional rather than relational interactions. This gap highlights the need for VAs that convey warmth or empathy to build trust over time (Mari et al., 2024). When users perceive a VA's voice as human-like due to its emotional expressiveness, trust increases as the interaction feels more familiar and relatable (Fakhimi et al., 2023). This research explores how human-like qualities in a VA's voice correlate with user trust.

H3. An enhanced perception of a synthetic voice as human fosters greater trust in the user.

The conceptual model (see Figure 3.1) illustrates the relationships between emotional inputs, emotional responses, human-like voice and trust in VAs.

Figure 3.1. Conceptual Model and Hypothesis



Notes: PE: Positive Emotions; NE: Negative Emotions; ER: Emotional Response; HLV: Human-Like Voice; TRU: Trust.

3.2.4. The moderating effect of voice type

The emotional tone in synthetic VA voices is pivotal in shaping ER. VAs that convey positive emotional cues effectively can evoke favorable reactions and build trust, whereas voices with flat or inappropriate tones may trigger negative responses, diminishing trust (Fakhimi et al., 2023). While VAs strive to balance task efficiency and user engagement, many still struggle to maintain emotional expressiveness, resulting in interactions that can feel mechanical or less authentic (Lucia-Palacios & Pérez-López, 2021).

Research highlights that voice elements like tone, pitch, timbre, and rhythm significantly impact emotional perception. High-pitched, fast-paced voices often convey excitement, while lower-pitched, slower voices are seen as calm or authoritative (Efthymiou et al., 2024; Scherer, 2003). Psychoacoustic studies reveal that listeners link vocal attributes with specific emotions, and these can be adjusted in voice synthesis to elicit desired responses (Laukka et al., 2005).

Although ER's role in human interactions is well-documented, less focus has been given to its impact on trust in AI systems like VAs (Zhang et al., 2021). Studies in human-computer and human-human communication suggest that variations in voice synthesis likely affect user perceptions and emotional responses to VAs. Torre et al. (2020) found that tone and expressiveness in virtual agents influenced perceptions of competence and trust. Fakhimi et al. (2023) similarly found that emotional cues in AI communication boost

perceived authenticity, crucial for trust.

Balanced emotional expressiveness is essential: overly enthusiastic or flat voices can seem inauthentic or indifferent, harming trust (Cheshin et al., 2018). VAs with warm, empathetic voices foster personal interactions, enhancing satisfaction and brand connection (Chin & Yi, 2022). Properly managed emotional cues help businesses create human-like interactions that resonate with users and drive positive outcomes (Cowan et al., 2015).

This study proposes that different VA voices—Siri, Alexa, and Google Assistant—modulate emotional expression, each evoking varied responses due to distinct voice synthesis designs (Laukka et al., 2005). It is hypothesized that conversational, expressive voices will elicit stronger positive ER and trust, while neutral voices may limit emotional connection. This will be empirically tested to see how different synthetic voices impact ER and trust.

Understanding how these voices influence ER is essential for designing VAs that build deeper trust and emotional resonance, especially where warmth or empathy enhances user experience.

H4. The transmission and impact of basic emotions vary based on the type of voice used by the VA.

3.2.5. Impact of Message Valence on Emotional Response and Trust

Positive-valence messages, like encouraging or reassuring statements, are expected to strengthen positive emotional responses when paired with a warm synthetic voice, particularly in customer service where users tend to trust VAs delivering positive content with warmth (Coker & Thakur, 2024). Studies show that positive message valence, combined with an engaging voice, enhances perceptions of warmth and empathy (Nass & Brave, 2005). Enthusiastic delivery of positive messages bolsters the VA's image as empathetic and reliable (Coker & Thakur, 2024), aligning with studies that found that voice integration in omnichannel experiences boosts brand trust and loyalty (Pagani et al., 2019). Studies have noted that positive interactions with AI voice assistants promote brand loyalty, positioning VAs as tools for sustained engagement (Huh, Kim, et al., 2023). Familiar voice characteristics also create moments of recognition, enhancing user enjoyment (Hall et al., 2024).

Conversely, negative-valence messages, such as critiques or warnings, tend to elicit negative responses, especially when delivered in a neutral or cold tone (Oya & Tanaka, 2023). Users are notably sensitive to negative messages, which can trigger reactions like frustration or anxiety, especially in contexts like healthcare or product returns (Cowan et al., 2015). This moderating role of message valence becomes crucial when VAs handle complex tasks like product recommendations or advice, where users' emotional responses to both message and tone impact trust and willingness to accept recommendations (Mari et al., 2024). Research shows positive messages often foster favorable attitudes toward recommendations, enhancing VA credibility (Beldad et al., 2016). Other studies have also emphasize that initial interactions with conversational AI build foundational trust, shaping user perceptions and decisions (Hildebrand & Bergner, 2021)

For instance, a positive product message can boost user confidence and purchasing likelihood, while a negative message may create distrust if the VA's tone fails to mitigate its impact (Schroeder & Epley, 2015). Exploring how message valence affects emotional responses in VA interactions will reveal whether tone amplifies or softens emotional content, aiding in the design of more empathetic digital assistants (Coker & Thakur, 2024). Accordingly, we propose the following hypothesis:

H5a. Positive basic emotions conveyed through positive-valence messages have a positive effect on Emotional Response than negative-valence messages.

H5b. Negative basic emotions conveyed through negative-valence messages have a negative effect on Emotional Response than positive-valence messages.

3.3. Methodology

3.3.1. Study design

This study employs an experimental design to examine the relationships between positive and negative emotional inputs, emotional response, human-like voice, and trust in virtual assistants, also considering the moderating effects of voice type and message valence. Participants were exposed to synthetic voices from widely used VAs (Siri, Google Assistant, and Alexa) to gauge their emotional responses.

Before launching the main experiment, a pilot study was conducted independently with 180 participants under the age of 50 who had previous experience with virtual assistants.

This preliminary phase served to validate the audio stimuli and test the attention and comprehension checks embedded throughout the survey. Participants were exposed to a sample of the audio clips, completed the full survey flow, and provided feedback on clarity, difficulty, and technical performance. Results from the pilot confirmed the effectiveness of the experimental manipulations and informed minor adjustments to the final study protocol.

Siri, Google Assistant, and Alexa were selected due to their prevalence in the virtual assistant market. As of 2023, they dominate the industry, powering billions of devices globally. Siri, integrated into over 1.5 billion Apple devices, is widely used by iPhone, iPad, and Mac users (Statista Staff, 2023). Google Assistant, accessible on over 2.5 billion devices, is present across smartphones, smart speakers, and smart TVs (Google, 2023). Alexa has a substantial share in the smart home market, particularly in the U.S., with over 70 million units sold worldwide (Amazon Staff, 2022). The inclusion of these VAs enables testing of emotional responses across highly recognized voice types, ensuring findings are applicable to common interactions with leading virtual assistants.

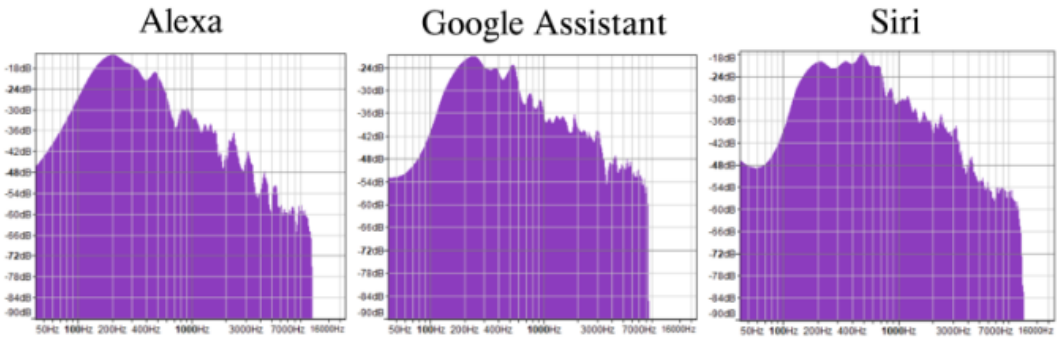
Participants listened to seven product reviews, either positive or negative, read by VAs—Siri, Alexa, and Google Assistant—in developer mode. Recorded as .wav files, each clip was about 10 seconds, simulating real-world applications where VAs deliver product recommendations that influence consumer trust and satisfaction (Mari et al., 2024).

A total of 42 unique stimuli were presented across versions, alternating between audio clips and follow-up questions on participants' experiences. Participants confirmed hearing each clip before continuing. To ensure attentiveness, they listened to a clip where the VA expressed a preference for disco music, with all participants correctly identifying the preference. Each audio file was uploaded to a custom web platform requiring full playback before survey responses. Security questions were included throughout to ensure attentive participation, with data flagged for review if questions were answered incorrectly.

Voice emotional impact varies based on qualities like whether it is whispered or shouted (Scherer, 2003). Differences in pitch, volume, and timbre significantly influence perceived emotional content, as these vocal cues shape emotion interpretation (Juslin & Laukka, 2003). To prevent bias, a spectral distribution analysis was performed for each voice using Audacity 3.5, confirming effective manipulation of pitch, timbre, and loudness. Figure 3.2 illustrates that the spectral distribution is similar across Alexa, Google, and Siri as they

deliver the same content, verifying that differences stem from voice type manipulations rather than other factors.

Figure 3.2. Spectral distribution of synthetic voices



After developing the preliminary study, AI and human-computer interaction experts reviewed the experimental design, participating in interviews to address potential issues and clarify uncertainties. Thirty volunteer users from the target demographic then completed the full study, providing feedback on the design and questionnaires. To verify the manipulation of voice type, participants rated the voice perception on a Likert scale from 1 (robotic/unpleasant) to 7 (human-like/pleasant) (Bartneck et al., 2009). ANOVA results confirmed significant differences across voices: $F = 5.692$, $p = 0.004$ for Robotic/Human and $F = 7.617$, $p = 0.001$ for Unpleasant/Pleasant. To ensure group homogeneity, perceived usefulness and ease of use, based on the Technology Acceptance Model (TAM), were applied as control variables. ANOVA results showed no significant differences in ease of use ($F = 0.409$, $p = 0.664$) or perceived usefulness ($F = 0.446$, $p = 0.640$), confirming that group differences were due to the experimental manipulation rather than pre-existing perceptions. To avoid potential bias from familiarity with the assistant's voice, participants were asked to identify which VA voice they were hearing, with only 34.2% correctly identifying it. Participants also indicated which assistant they typically used. A chi-square test confirmed no dependent relationship between the type of VA typically used and the assistant exposed in the study (chi-square: 13.291, $p = 0.205$).

3.3.2. Sample

Each participant was randomly assigned to one version, hearing clips from only one assistant delivering either positive or negative messages. This randomization ensured an equal distribution across six conditions, allowing for a balanced comparison of how voice

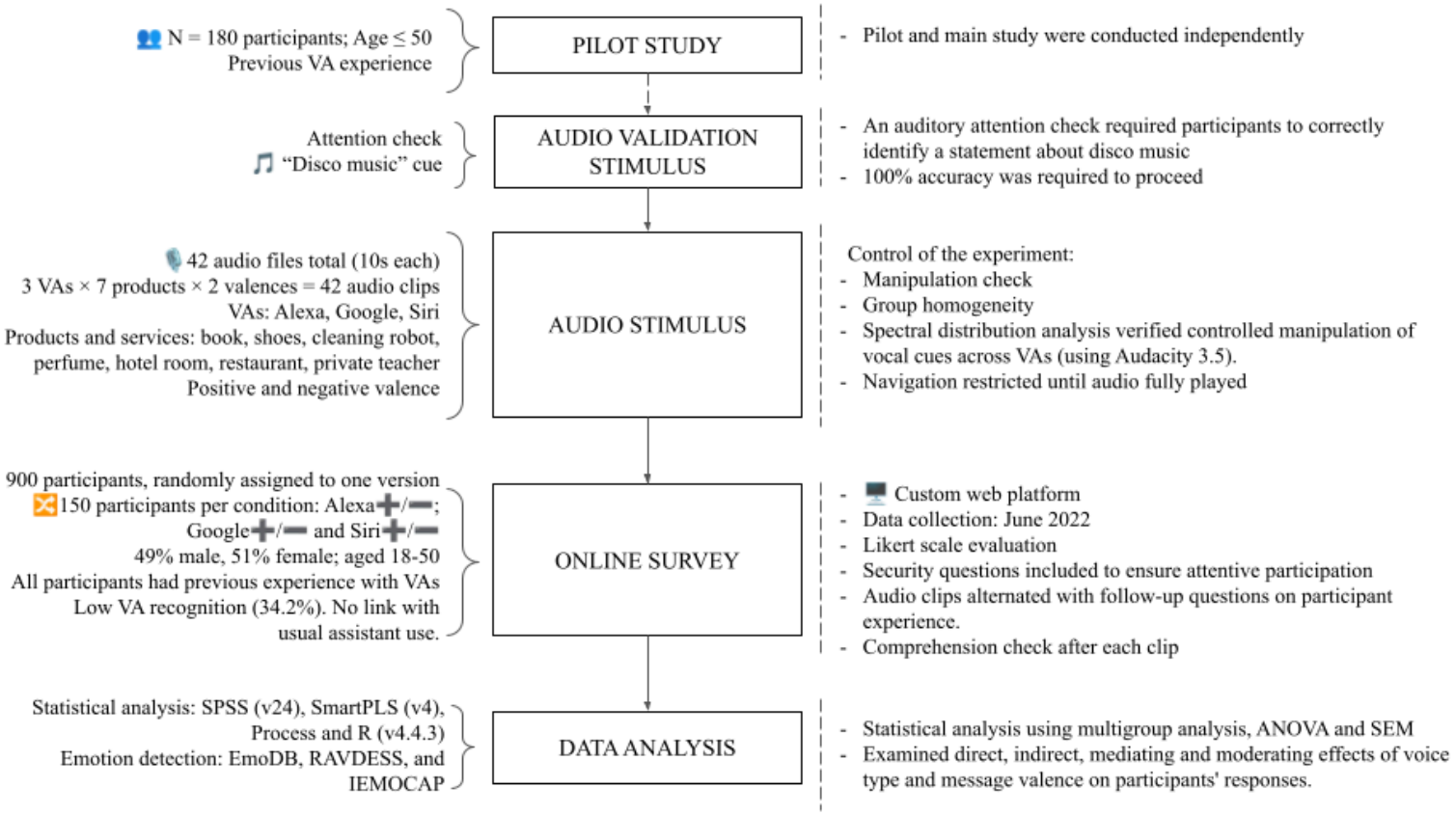
type and message valence influenced emotional response and trust, with a consistent questionnaire across all versions.

A total of 900 participants were recruited, nearly evenly split by gender (49% male, 51% female), aged 18-50 (average age 35.19). All confirmed prior experience with virtual assistants, ensuring familiarity with the technology. This age group aligns with data indicating that individuals aged 18-50 account for over 70% of VA interactions, as they are primary users of voice-activated devices (Voicebot.ai, 2023).

The survey took place in June 2022 on an online platform provided by a market research firm. Participants were randomly assigned to one of six questionnaire versions, each identical in structure but differing by virtual assistant (Siri, Alexa, or Google Assistant) and message valence (positive or negative). Each condition—Siri with positive and negative messages, Google Assistant with positive and negative messages, and Alexa with positive and negative messages—had 150 participants, totaling 900. Including both positive and negative messages provided a comprehensive view of typical user interactions with virtual assistants (Fakhimi et al., 2023).

Figure 3.3 provides an overview of the experimental design and data collection procedure used in this study. The flowchart outlines the sequence of key stages, including the pilot study, stimulus validation, audio exposure, online survey, and data analysis. It summarizes how participants were recruited and exposed to different voice assistant conditions, how the stimuli were developed and validated, and how the data were collected and analyzed.

Figure 3.3. Data collection procedure



3.3.3. Measuring variables

This study examined key variables related to ER, trust, and VA perception, focusing on valence and arousal as primary emotional dimensions, with trust measured based on participants' perceptions of each assistant after exposure to audio stimuli.

Positive and negative emotions

This study's decision to measure basic emotions follows a well-established approach in psychological and human-computer interaction research. Basic emotions, as proposed by Ekman (1992), are seen as universal across cultures and essential for understanding emotional processing in both interpersonal and human-computer interactions. Research has shown that assessing basic emotions can offer key insights into user experiences with VAs. For instance, Fakhimi et al. (2023) found that basic emotions elicited in VA interactions significantly affected user trust and engagement. Similarly, Schroeder & Epley (2015) demonstrated that positive emotions fostered trust and rapport in digital interactions, while

negative emotions led to disengagement. This approach is also reflected in studies on synthetic voices, such as (Cuciniello et al., 2023), where participants' emotional responses to synthetic voices were assessed along basic emotional dimensions to evaluate trust and satisfaction.

The use of basic emotions in this study aligns with established methodologies in psychological and HCI research, providing a solid framework for analyzing how VA voices influence emotional responses and trust. Focusing on basic emotions enables a clear and measurable examination of user experiences, offering valuable insights into the emotional dynamics of human-computer interaction.

Following the literature, participants' basic emotions were measured using 15 distinct emotions (Table 3.8, Appendix), including 6 positive emotions (PE) and 9 negative emotions (NE) (Ekman, 1992). These emotions were selected to capture a broad range of reactions, from moderate to extreme states. Participants rated the extent to which each emotion was evoked by the VA's voice and message content on a 7-point Likert scale, from strong disagreement (1) to strong agreement (7).

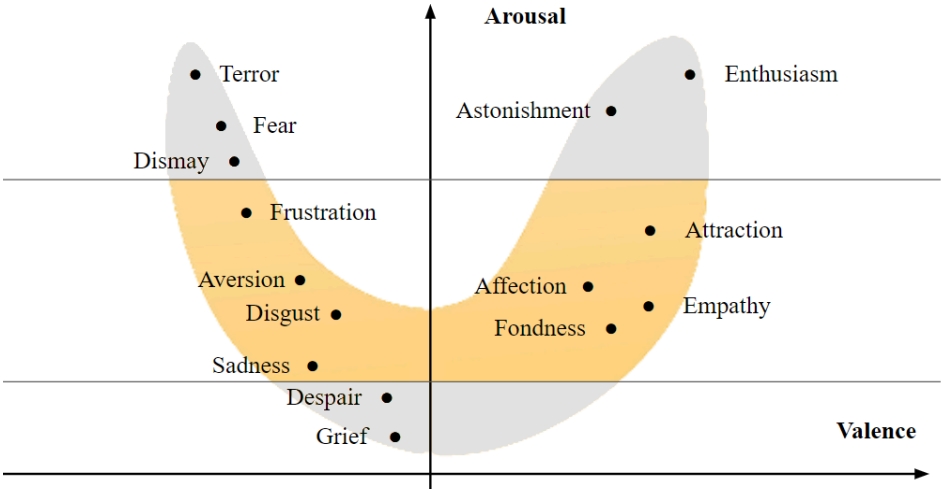
An exploratory factor analysis on the collected responses showed that extreme emotions (e.g., extreme joy, despair) did not load as consistently as moderate emotions, suggesting that extreme states may not be as effective in fostering human-like voice perceptions for VAs. This finding aligns with previous research, indicating that extreme emotional responses can lead to discomfort or mistrust (Huang & Sénécal, 2023); moderate expressions were more positively associated with trust perceptions.

Four positive emotions—empathy (0.838), affection (0.930), fondness (0.930), and attraction (0.857)—demonstrated acceptable reliability, with factor loadings and Cronbach's alpha values above 0.7, confirming their strong contribution to the positive emotional response factor (Nunnally & Bernstein, 1994). Similarly, four negative emotions—aversion (0.753), disgust (0.859), frustration (0.896), and sadness (0.909)—showed reliability with factor loadings above 0.7, supporting their association with the negative emotions factor. Positive emotions like enthusiasm and astonishment, and negative emotions like dismay, despair, fear, and terror, were excluded from the analysis.

Comparing our results with previous studies, Figure 3.4 shows the distribution of basic emotions according to valence and arousal based on Russell (1980) circumplex model

of affect. The emotions highlighted in yellow correspond successfully with those included in the analysis, characterized by higher valence and moderate arousal. In contrast, the gray-shaded emotions did not load, reflecting their extreme nature in either low valence or high arousal.

Figure 3.4. Circumplex model of affect



Notes: Adapted from (Russell, 1980).

Emotional Response, Human-Like and Trust

To measure ER, valence was assessed with the question, "The voice of this assistant seemed to me...," rated on a 7-point Likert scale from 1 ("Very negative") to 7 ("Very positive"), following HCI standards where valence reflects the positivity or negativity of AI-driven user experiences (Kim et al., 2021). Similarly, arousal was measured by asking, "The voice of this assistant made me feel...," with options from 1 ("Very relaxed") to 7 ("Very nervous"), capturing the emotional intensity evoked by the synthetic voice and its effect on engagement and trust (Bradley & Lang, 1994; Kim et al., 2021).

For the Human-like Voice construct, the scale from Patrizi et al. (2021) was used to assess perceived human-likeness of the synthetic voices, rated on a 7-point Likert scale from 1 = Strongly Disagree to 7 = Strongly Agree. For Trust, the scale from Charalambous et al. (2016) gauged participants' trust in the VAs during interactions, with responses also on a 7-point Likert scale from 1 = Strongly Disagree to 7 = Strongly Agree, providing an in-depth view of trust toward the VAs after hearing their voices. Participants were consistently reminded that the questions focused on the voice of the assistant.

3.4. Analyses and results

3.4.1. Analysis I: Model estimation

Reliability analysis

To verify the scales' dimensionality, reliability, and validity, we undertook both exploratory and confirmatory factor analyses using SPSS (version 24) and partial least squares (PLS) with statistical software SmartPLS 4. These analytical results are detailed in Table 3.1. Discriminant validity was studied in Table 3.2 using established criteria (Fornell & Larcker, 1981; Hair, 2009). Furthermore, we conducted an in-depth examination of the heterotrait–monotrait (HTMT) ratio values (Table 3.3), following the guidance of Henseler et al. (2016).

Table 3.1. Measurement model results

Variable	Item	Mean	SD	λ	t-value	CA	CR	AVE
PE	PEI	5.935	0.780	0.924	160.090***	0.913	0.914	0.794
	PEI	5.211	1.024	0.867	85.116***			
	PEI	5.074	1.049	0.916	141.734***			
	PEI	5.704	0.788	0.854	76.983***			
NE	NEI	1.981	1.542	0.827	48.047***	0.919	0.936	0.805
	NEI	1.647	1.332	0.940	189.559***			
	NEI	1.562	1.513	0.939	148.822***			
	NEI	1.513	1.247	0.878	71.737***			
ER	ER1	6.254	0.698	0.888	101.848***	0.745	0.746	0.797
	ER2	6.233	0.660	0.898	123.619***			
HLV	HLV1	5.112	1.035	0.922	120.872***	0.924	0.925	0.869
	HLV2	5.159	1.067	0.943	191.976***			
	HLV3	5.267	1.006	0.932	188.247***			
TRU	TRU1	5.409	0.876	0.921	149.322***	0.905	0.907	0.841
	TRU2	5.380	0.934	0.928	162.385***			
	TRU3	5.510	0.872	0.902	96.086***			

Notes: CA: Cronbach's alpha; CR: composite reliability; AVE: average variance extracted. ***sig. < 0.01. PE: Positive Emotions; NE: Negative Emotions; ER: Emotional Response; HLV: Human-like Voice; TRU: Trust.

Table 3.2. Correlations among factors

	PEI	NEI	TRU	HLV	ER
PEI	0.891				
NEI	-0.166	0.897			
TRU	0.629	-0.188	0.917		
HLV	0.605	-0.116	0.790	0.932	
ER	0.634	-0.362	0.499	0.456	0.893

Notes: PE: Positive Emotions; NE: Negative Emotions; ER: Emotional Response; HLV: Human-like Voice; TRU: Trust.

Table 3.3. Heterotrait-monotrait ratio (HTMT)

	PEI	NEI	TRU	HLV	ER
PEI					
NEI	0.176				
TRU	0.690	0.204			
HLV	0.658	0.124	0.863		
ER	0.769	0.433	0.605	0.548	

Notes: PE: Positive Emotions; NE: Negative Emotions; ER: Emotional Response; HLV: Human-like Voice; TRU: Trust.

Structural model analysis

We assessed the structural model's efficacy through path analysis, employing a bootstrapping approach with 5,000 resamples. A crucial part of our evaluation of the PLS-SEM model involved analyzing the effect size (f^2) to understand the importance of the predictors in the model (Table 3.4). Additionally, we evaluated the predictive capacity of our structural equation model using the Stone–Geisser Q^2 predictive relevance index. These results are shown in Table 3.5.

Table 3.4. Effect size f²

Predictor	Endogenous construct	Effect size f²
Positive Emotion Indicators	Affective Arousal	0.641
Negative Emotion Indicators	Affective Arousal	0.128
Human-like Voice	Trust	1.659
Emotional Response	Human-like Voice	0.263

Notes: An f² value of 0.02, 0.15, and 0.35 is interpreted as a small, medium, and large effect, respectively, according to (Cohen, 1988).

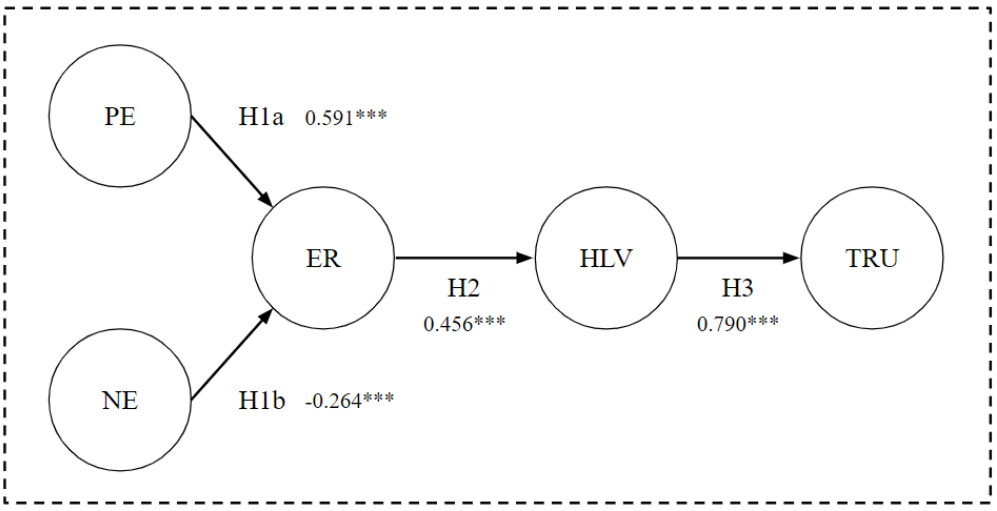
Table 3.5. Predictive index Q²

Endogenous construct	Q2 Predict Value	Predictive Relevance
Trust	0.242	Moderate
Human-like voice	0.255	Moderate
Emotional Response	0.467	Substantial

Notes: According to the criteria established by Chin (1998), these values, being greater than zero, indicate moderate to substantial predictive relevance for the constructs within the model.

The proposed hypotheses were all supported (Figure 3.5), and the model accounted for 62.4% of the variance in trust.

Figure 3.5. Estimated model



***p < .001. Notes: Estimated model. Notes: PE: Positive Emotions; NE: Negative Emotions; ER: Emotional Response; HLV: Human-like Voice; TRU: Trust.

3.4.2. Analysis II: Moderating effect of voice type

As voice technology advances, VAs are increasingly expected to convey emotions that feel natural and human-like, enhancing user trust. H4 posits that the type of synthetic voice used by the VA (Siri, Alexa, or Google Assistant) moderates the relationship between ER and HLV. While emotional responses to VAs are essential to user interactions, the ability of different voices to mediate these responses and influence perceptions of human-likeness is underexplored.

To test this hypothesis, a multigroup bootstrap analysis was conducted in SmartPLS, allowing us to examine how voice type moderates the relationship between ER (valence and arousal) and perceived HLV. The analysis revealed significant differences in how Siri, Alexa, and Google Assistant convey emotions, which in turn influences users' perception of their voices' human-likeness.

Results show that Siri's voice is most effective in eliciting strong ER, showing significant differences with Google across all relationships and with Alexa in all but one. Siri's positive emotions have a stronger impact on ER than those of the other two assistants, and its negative emotions have a lower impact than Alexa's. Google appears less effective in conveying negative emotions, while Alexa, though capable of conveying both, shows a consistently greater impact of negative emotions, leading to less favorable outcomes. Siri's ER most strongly influences HLV, ultimately enhancing trust in the device. Full results are presented in Table 3.6.

Table 3.6. Multigroup analysis

	Total β (t-value)		
PE → ER	0.591 (25.448***)		
NE → ER	-0.264 (9.942***)		
ER → HLV	0.456 (15.486***)		
HLV → TRU	0.790 (52.928***)		
	β (t-value)	**β (t-value)**	**Difference β (t-value)**
	Alexa	Google	
(PE → ER)	0.474 (10.074***)	0.334 (5.222***)	0.140 (4.852)
(NE → ER)	-0.378 (9.245***)	-0.099 (0.984)	-0.279 (8.261***)
(ER → HLV)	0.366 (6.723***)	0.248 (3.278***)	0.118 (3.445)

(HLV → TRU)	0.771 (27.636***)	0.712 (20.609***)	0.059 (7.027)
	β (t-value)	**β (t-value)**	**Difference β (t-value)**
	Siri	Alexa	
(PE → ER)	0.694 (20.855***)	0.474 (10.074***)	0.22 (10.593***)
(NE → ER)	-0.273 (6.544***)	-0.378 (9.245***)	0.105 (-1.997**)
(ER → HLV)	0.487 (10.296***)	0.366 (6.723***)	0.121 (3.583**)
(HLV → TRU)	0.800 (28.659***)	0.771 (27.636***)	0.029 (1.023)
	β (t-value)	**β (t-value)**	**Difference β (t-value)**
	Google	Siri	
(PE → ER)	0.334 (5.222***)	0.694 (20.855***)	-0.360 (-15.633***)
(NE → ER)	-0.099 (0.984)	-0.273 (6.544***)	0.174 (-5.56***)
(ER → HLV)	0.248 (3.278***)	0.487 (10.296***)	-0.239 (-7.018***)
(HLV → TRU)	0.712 (20.609***)	0.800 (28.659***)	-0.088 (-8.05***)

Notes: PE: Positive Emotions ; NE: Negative Emotions; ER: Emotional Response; HLV: Human-like Voice; TRU: Trust. Significant coefficients at level *** <0.01.

3.4.3. Analysis III: Automatic emotion detection

To reinforce our findings, we conducted Automatic Emotion Detection (AED) using machine learning techniques to analyze emotional content in VA voices. This objective analysis aimed to classify emotions like happiness, sadness, anger, and neutrality in synthetic voices, extending our initial user-centered assessments. Unlike previous analyses based on participants' subjective perceptions, AED allowed us to capture emotional nuances conveyed by VA voices, providing an additional validation layer, particularly highlighting Siri's effectiveness in eliciting robust ER outcomes.

In human-computer interaction, AED integrates signal processing and machine learning to classify emotions in voice data, widely applied in customer service, mental health, and education for analyzing tone, intensity, rhythm, and pitch (Sezgin et al., 2012). For our analysis, temporal and spectral parameters—such as rhythm, pitch, energy distribution, and Mel-frequency cepstral coefficients (MFCC)—were extracted (Jacob, 2016). Support Vector Machines (SVM) were used to classify emotions based on these features (Kudiri et al., 2012).

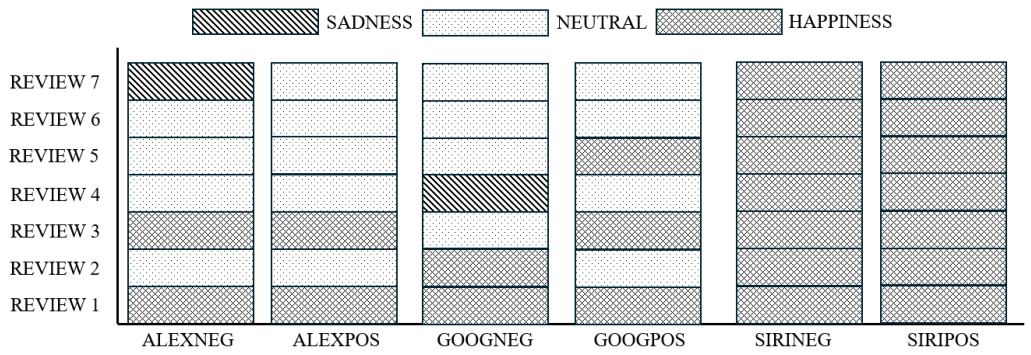
To ensure model accuracy, we combined datasets (EmoDB, RAVDESS, and IEMOCAP), providing diverse emotional contexts and enhancing the model's detection

capability across VAs. Feature standardization, like Z-score normalization, ensured comparability across datasets, reducing bias (Ghaleb et al., 2019). By applying cross-corpus training (Pastor et al., 2022), the model was trained across various datasets, improving classification accuracy.

The SVM model detected four key emotions—happiness, sadness, anger, and neutral—from audio clips of Siri, Alexa, and Google Assistant. Results showed that Siri's voice consistently conveyed happiness, aligning with positive emotional cues, while Google Assistant and Alexa displayed a broader range of emotional classifications, including negative and neutral emotions. These findings support previous analyses, indicating Siri's stronger expressiveness in evoking positive responses. Figure 3.6 illustrates the emotional classifications across assistants, highlighting differences in emotional transmission.

This AED analysis confirms earlier conclusions, showing significant variation in VAs' ability to convey specific emotions. By integrating AED, this study provides objective evidence of synthetic voice impacts on ER, enhancing our understanding of VA design for eliciting appropriate emotional reactions, which in turn foster trust.

Figure 3.6. Classification of automatic emotion detection



3.4.4. Analysis IV: Moderating effect of message valence

The emotional content of messages conveyed by VAs significantly shapes users' emotional responses and perceptions of trust. We hypothesized that message valence—whether positive or negative—moderates the relationship between ER and users' emotional reactions. Specifically, H5a suggests that positive-valence messages lead to more favorable emotional responses, while H5b proposes that negative-valence messages result in less favorable responses.

To test this, participants were randomly assigned to listen to either positive or negative audio clips, all structured as product reviews delivered by VAs. Each participant was exposed to one set of valence-specific clips, and their emotional reactions were recorded. The messages consisted of positive and negative product reviews, selected based on valence scores from the Affective Norms for English Words (ANEW) database (Bradley & Lang, 1999) and the NRC Emotion Lexicon (Mohammad & Turney, 2013). A manipulation check confirmed that participants accurately perceived the messages' valence. Participants rated the content on a scale from 1 (very negative) to 7 (very positive). An independent-samples t-test showed that positive messages received a mean score of 5.23, while negative messages scored 2.47 ($t = 865.475$; $p < 0.001$), validating the valence classification.

As shown in Table 3.7, the results support H5a, demonstrating that positive-valence messages have a significantly stronger effect on ER than negative-valence messages. This finding suggests that positive messages are more effective in fostering favorable ER outcomes, underscoring the importance of aligning message content with positive emotional cues to enhance user satisfaction. These outcomes highlight the differential impact of message valence on ER, indicating that positive messages foster a stronger emotional connection between the user and the VA.

Table 3.7. Differences among positive and negative messages

	β (t-value)	**β (t-value)**	**Difference β (t-value)**
General	Positive messages	Negative messages	
(PE → ER)	0.662 (22.370***)	0.439 (10.049***)	0.223 (***)
(NE → ER)	-0.200 (5.050***)	-0.170 (4.411***)	-0.03

Notes: PE: Positive Emotions; NE: Negative Emotions; ER: Emotional Response. Significant coefficients at level *** <0.01.

However, H5b was not supported, as no significant differences were found in the emotional responses to negative-valence messages. The analysis revealed that negative emotions had a similar impact on emotional response, regardless of whether the message was perceived as positive or negative.

3.5. Conclusions and discussion

3.5.1. Key findings

This study examined how ER, shaped by the synthetic voices of VAs, affects user perception of human-likeness and trust. Our analyses provided critical insights into ER and message valence's role in shaping VA perceptions, aligning with and expanding existing literature.

The primary contribution lies in demonstrating that VA voices carry emotional weight, capable of conveying basic emotions that influence ER, thereby enhancing perceptions of human-likeness and trust. By focusing on popular VAs like Siri, Alexa, and Google Assistant, and using messages of varying valence, this study confirms that synthetic voices can transmit emotional cues effectively, making VAs seem more human-like. This emotional expressiveness not only strengthens user trust but also enhances interactivity, as users respond more positively to VAs that elicit favorable emotions. Notably, Siri's voice proved most effective in eliciting strong ER, suggesting that Siri's voice synthesis may be particularly attuned to conveying emotional nuances.

Analysis I supported H1a and H1b, demonstrating that VAs can convey both positive and negative basic emotions—such as empathy, affection, attraction, aversion, disgust, frustration, and sadness—which affect users' ER differently. Positive emotional expressions in VA voices significantly enhanced ER, increasing user satisfaction, aligning with previous findings that positive emotional cues in technology interactions boost satisfaction (Schroeder & Epley, 2015; Shin et al., 2021) and build trust by creating warmth and reliability (Schroeder & Epley, 2015; Xiong et al., 2024). In contrast, negative emotions were shown to reduce ER, though some studies indicate that conveying mild negative emotions can be valuable in realistic or empathetic responses (Xiong et al., 2024). However, extreme negative emotions, like terror, may lead to user discomfort, risking the perception of an overly artificial or dehumanized assistant. By evaluating a range of basic emotions, this study highlights that moderate, balanced emotional expressions—aligned with Flow Theory—are most effective in fostering positive responses, while extremes may undermine trust.

H2 was confirmed, indicating that heightened ER reinforced users' perception of the VA's human-like voice qualities. This aligns with Fakhimi et al. (2023), who found that emotional expressiveness in artificial agents enhances human-likeness and solidifies trust.

Similarly, Calahorra-Candao & Martín-de Hoyos (2024) showed that anthropomorphic qualities in VA voices positively impacted perceived safety in voice shopping, closely tied to trust. These findings underscore that moderate, positive emotional expressiveness is more effective in fostering trust and building a human-like connection, while extreme emotions can detract from trust, aligning with Flow Theory's balance concept (Csikszentmihalyi, 1997), which suggests users achieve optimal satisfaction with moderate emotional experiences.

H3 was also supported, showing that a stronger perception of a VA's human-likeness directly enhances trust. This aligns with Chérif & Lemoine (2019), which found that more human-like voice qualities in AI systems create empathy and relatability, essential for building trust in sustained interactions. This study extends the Computational Model of Trust by integrating ER into the process and purpose dimensions, suggesting that emotionally attuned synthetic voices foster more engaging, human-like interactions, ultimately enhancing user satisfaction.

Analysis II tested H4, showing that the type of VA voice (Siri, Google Assistant, Alexa) significantly moderated the relationship between ER, human-likeness perception (HLV), and trust. Siri consistently evoked stronger positive emotional responses, aligning with prior findings that Siri's voice synthesis feels more natural and emotionally engaging (Fakhimi et al., 2023). In contrast, Google Assistant exhibited a broader emotional range but with less impact on trust, consistent with (Schroeder & Epley, 2015), who noted that variations in voice synthesis affect trust across VAs. Emotion detection analysis further supported these results, as Siri's voice was consistently associated with happiness, highlighting Siri's effectiveness in conveying positive emotions and enhancing ER and trust. By examining how voice type moderates the ER-HLV relationship, this study underscores the importance of selecting synthetic voice characteristics to boost emotional trust and human-likeness, supporting continuous refinement in voice technology for trust-based user relationships across domains.

This study focuses on the emotional impact of synthetic voices, examining which voice types are most effective for evoking responses essential to building trust. Although it doesn't specifically dissect tone, intonation, or pacing, it assesses how users react emotionally to the overall voice characteristics of various VAs. Findings show that Siri's voice consistently generates stronger positive responses, indicating it is perceived as more approachable and engaging. This enhanced connection likely stems from Siri's specific vocal

qualities, which convey warmth and empathy, thus reinforcing user trust. Such differences highlight the importance of designing VA voices thoughtfully to enhance user trust and satisfaction. Specifically, the study demonstrates that variations in emotional expressiveness across VA voices significantly shape perceptions of the VA's human-likeness and reliability, which are crucial for fostering positive, trust-based interactions.

Lastly, Analysis IV tested H5a and H5b, examining the moderating effect of message valence. The results supported H5a, showing that positive-valence messages had a more favorable impact on users' ER. However, H5b was not supported, as negative-valence messages did not significantly differ in their effect on ER, regardless of content. The results align with findings suggesting that positive emotional cues have a favorable impact on user engagement, while negative cues tend to encourage disengagement in AI interactions (Shin et al., 2021; Zhang et al., 2021).

These findings underscore the importance of both the emotional expressiveness of VA voices and the valence of conveyed messages in shaping user trust. By testing a broad range of basic emotions and confirming the role of moderate, positive emotional expressiveness, this study addresses a gap in the literature and offers new insights into VA design. These results build on literature emphasizing the importance of emotionally intelligent AI systems that foster deeper, more human-like interactions through expressive voices and appropriate message content (Shin et al., 2021; Zhang et al., 2021). By designing VAs that transcend functional efficiency to evoke rich emotional experiences, developers can enhance user satisfaction and promote long-term trust.

3.5.2. Implications for theory

This study provides valuable insights into Human-Computer Interaction and VA design, specifically on how emotional response influences trust and human-likeness in AI systems. By integrating the Computational Model of Trust (Lee & See, 2004), it demonstrates that trust in human-computer interactions depends not only on VA performance but also on the emotional expressiveness of the interaction. This research extends the model by positioning ER as a core element influencing trust's process and purpose dimensions, showing that emotionally responsive synthetic voices create a more human-like and engaging experience aligned with user expectations and values.

First, this study advances emotional response theory by emphasizing the importance of valence and arousal in interactions with synthetic voices. While prior research explored ER in human contexts (Ekman, 1992; Russell & Barrett, 1999), this study extends these principles to VA interactions, specifically their voices and the trust they inspire. By validating the role of positive valence and moderate arousal in fostering trust, the study highlights the importance of emotionally expressive VAs for enhancing user engagement (Xiong et al., 2024). The Computational Model of Trust provides a structured framework for understanding ER's role in trust-building, expanding beyond task performance to include emotional dynamics.

The use of valence and arousal is supported by prior research, which shows their efficacy in assessing emotional responses in voice-based interactions. Positive valence correlates with user satisfaction and trust, while moderate arousal fosters engagement without overwhelming users (Schroeder & Epley, 2015). Conversely, extremes in either dimension can disrupt trust (Schroeder & Epley, 2015), emphasizing the need for balanced emotional expressiveness in VAs.

Second, the study bridges emotional response theory with anthropomorphism research by demonstrating how human-like voice qualities affect perceptions of VA reliability. Traditionally studied in robotics (Schroeder & Epley, 2015), anthropomorphism is applied here to synthetic voices, revealing that emotional expressiveness enhances human-likeness. This aligns with findings linking emotional expressiveness to perceived authenticity in virtual agents (Fakhimi et al., 2023), reinforcing the idea that emotionally expressive voices positively influence trust.

The study further demonstrates how synthetic voices translate emotional expressions into trust, showing that ER impacts perceptions of human-likeness, which then influence trust. By examining how Siri, Google Assistant, and Alexa mediate this relationship, the research adds depth to emotional intelligence studies in AI, supporting findings that voice synthesis variations create different user experiences (Fakhimi et al., 2023). The Computational Model of Trust supports this perspective, explaining how voice characteristics shape user perceptions and validating the model's relevance in VA research.

This aligns with findings on the acceptance of hedonic information systems, which highlight the importance of user experience and emotional engagement as key factors driving

system adoption (Van der Heijden, 2004). Virtual assistants designed with emotionally expressive and user-friendly voices are more likely to be embraced by users, as they provide both functional utility and pleasurable interactions.

This research extends the body of knowledge by shifting focus from traditional areas like visual design or conversational style to the largely overlooked yet impactful dimension of voice characteristics. By doing so, it contributes a novel perspective to understanding how emotional expressiveness in synthetic voices shapes user trust and perceived humanity.

This study also supports Flow Theory, demonstrating how emotional dynamics in VA interactions can facilitate or disrupt flow experiences. Flow Theory posits that optimal user satisfaction occurs when challenge and skill are balanced (Csikszentmihalyi, 1997). Positive valence and moderate arousal enhance user immersion and satisfaction, while extreme emotional expressions disrupt flow. This application highlights the importance of calibrated emotional expressiveness in AI to ensure balanced experiences.

Finally, the study introduces the moderating effect of message valence as a new dimension in AI-mediated communication research. While prior studies focused on voice tone (Schroeder & Epley, 2015), this study shows that message content—whether positive or negative—significantly impacts ER. Positive-valence messages evoke stronger positive emotions than negative ones, demonstrating how content interacts with voice to shape trust. The Computational Model of Trust supports this, suggesting that aligning content with user expectations and emotions is critical for trust-building.

In conclusion, this study enriches theories of emotional response, Flow Theory, and human-likeness in VA contexts, offering valuable insights into how synthetic voices can enhance user trust. By examining the roles of voice type and message valence, this research contributes to the AI-human interaction literature and opens new avenues for designing emotionally intelligent VAs.

3.5.3. Implications for practice

These findings highlight the opportunity for businesses to strategically design virtual assistants with voices that reflect and reinforce their brand identity. Emotional branding, often achieved through sensory cues such as signature scents, has proven to be a powerful tool in creating lasting consumer connections (Pagani et al., 2019). In a similar way, a carefully

designed synthetic voice can serve as an auditory representation of the brand, creating a unique and memorable experience for the user. A warm, empathetic, and consistent voice, tailored to convey positive emotions, not only enhances trust and satisfaction but also builds stronger associations between the consumer and the brand (Gefen & Straub, 2004).

For companies using VAs in customer service, e-commerce, or healthcare, message valence—whether content is positive or negative—significantly impacts user trust and satisfaction. VAs should be programmed to provide positive, encouraging responses, as this approach strengthens customer satisfaction and loyalty. Emotionally expressive voices can amplify positive messages and soften negative ones, promoting a balanced user experience aligned with emotional expectations (Cuddy et al., 2011). Additionally, context-aware messaging that adjusts tone based on user emotions can further personalize interactions. For marketers, emotionally expressive VAs can improve brand perception and foster trust. In fields like e-commerce, where users increasingly rely on VAs for purchasing decisions, emotionally engaging interactions provide a competitive advantage. Investing in emotionally intelligent VAs can help create a personable experience, building trust in both the technology and brand.

This study offers valuable insights for VA designers, developers, and marketers aiming to build user trust through emotional expressiveness. First, developers can leverage findings on ER—valence and arousal—to craft more engaging, trustworthy assistants. Positive cues in synthetic voices, such as warmth, empathy, and moderate enthusiasm, enhance perceptions of human-likeness and reliability. Developers should focus on balancing positive valence with moderate arousal to foster comfort and trust, refining voice synthesis to deliver subtle, human-like emotional nuances suitable for context.

Siri has been shown to effectively transmit positive emotions, eliciting feelings of trust and human-likeness that foster stronger consumer relationships (Hall et al., 2024; Schroeder & Epley, 2015). By adopting a similar approach, businesses can design voices that align with their brand's personality—whether it be calm and reassuring for healthcare contexts or upbeat and engaging for retail environments. These voices could function as a critical element of the brand's identity, ensuring that whenever consumers hear the voice, they immediately associate it with the company, much like recognizing a jingle or logo.

Emotional expressiveness also enhances perceived competence and authenticity.

Users are more likely to trust VAs that show appropriate emotional responses, signaling understanding and attentiveness. Context-adaptive emotional expressions significantly boost user satisfaction and trust (Paradeda et al., 2024). However, over-expressiveness or unsuitable cues can harm trust, as users may see the assistant as inauthentic (Cheshin et al., 2018). Balancing functionality with emotional expressiveness is crucial for effective and trusted interactions (McLean & Osei-Frimpong, 2019).

Finally, VA designers should recognize that extremes in emotion—whether overly enthusiastic or too flat—can reduce trust. Emotional balance is essential, particularly in business and healthcare, where professionalism and clarity are expected. Siri's balanced emotional tone effectively evokes positive responses without extremes, enhancing trust by appearing more natural. Ensuring VAs convey context-appropriate emotional tones is key to maintaining trust and reliability.

In summary, this study provides actionable guidance for designing emotionally intelligent VAs that resonate with users and support positive experiences. By strategically incorporating emotional cues into voice and message design, developers and businesses can foster deeper, meaningful interactions across diverse applications.

3.5.4. Limitations and future research

While this study offers valuable insights into how ER—comprising valence and arousal—affects user trust in VAs, several limitations should guide future research. First, this research focused on emotional responses generated by VA voices without examining specific vocal attributes like pitch, tone, or flow, which are critical in voice perception and may impact users' emotional connections with and trust in VAs. Future studies could explore these vocal properties more deeply to understand their role in building trust. Additionally, examining voice modulation and its dynamic effects during prolonged interactions could provide a more nuanced understanding of VA perception over time.

Second, this study was limited to Siri, Google Assistant, and Alexa. Future research could include other VAs with varied linguistic capabilities or cultural contexts to assess the generalizability of the findings. Third, while this study measured emotional responses and trust at specific interaction moments, longitudinal research could track user trust over multiple interactions, offering insights into the sustainability of trust as VAs become more

embedded in daily life.

Finally, this study did not consider demographic factors like age or cultural background, which may influence emotional responses. Future studies could investigate how these variables, along with differences in gender or personality traits, shape user perceptions. Cross-cultural studies would help determine if these findings hold across regions and cultural contexts. Building on the systematic review by Lee et al. (2023) future research should also explore the organizational implications of implementing emotionally intelligent VAs, focusing on how they influence long-term customer trust and operational efficiency across industries.

Future research could build on the work of Luo et al. (2023) to further explore how emotion-regulatory strategies in VAs might optimize customer service and strengthen brand identity in diverse business applications. Specifically, future studies could examine how VAs with advanced emotion-regulation capabilities adapt to different user profiles and contexts, such as managing customer dissatisfaction in e-commerce or enhancing user engagement in healthcare. Investigating how these strategies impact long-term customer loyalty, brand perception, and organizational outcomes would provide deeper insights into their effectiveness.

In conclusion, while this study establishes a foundation for understanding the relationship between emotional response and trust in VAs, future research should explore additional vocal attributes, analyze real-time interactions, and consider the long-term effects of VA use to further develop these systems' emotional intelligence.

3.6. References

Amazon Staff. (2022, noviembre 30). *Amazon announces its biggest holiday shopping weekend ever: Top sellers included the Nintendo Switch and the Echo Dot*. US About Amazon. https://www.aboutamazon.com/news/retail/amazon-best-sellers-black-friday-cyber-monday

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, *1*, 71-81.

Beldad, A., Hegner, S., & Hoppen, J. (2016). The effect of virtual sales agent (VSA) gender–product gender congruence on product advice credibility, trust in VSA and online vendor, and purchase intention. *Computers in human behavior*, *60*, 62-72.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, *25*(1), 49-59.

Bradley, M. M., & Lang, P. J. (1999). Affective Norms for English Words (ANEW): Technical manual and affective ratings. *The Center for Research in Psychophysiology, University of Florida*.

Brave, S., & Nass, C. (2007). Emotion in human-computer interaction. En *The human-computer interaction handbook* (pp. 103-118). CRC Press.

Calahorra-Candao, G., & Martín-de Hoyos, M. J. (2024). The effect of anthropomorphism of virtual voice assistants on perceived safety as an antecedent to voice shopping. *Computers in Human Behavior*, *153*, 108124.

Casado-Aranda, L.-A., Dimoka, A., & Sánchez-Fernández, J. (2019). Consumer processing of online trust signals: A neuroimaging study. *Journal of Interactive Marketing*, *47*(1),

159-180.

Charalambous, G., Fletcher, S., & Webb, P. (2016). The Development of a Scale to Evaluate Trust in Industrial Human-robot Collaboration. *International Journal of Social Robotics*, *8*(2), 193-209. https://doi.org/10.1007/s12369-015-0333-8

Chérif, E., & Lemoine, J.-F. (2019). Anthropomorphic virtual assistants and the reactions of Internet users: An experiment on the assistant's voice. *Recherche et Applications En Marketing (English Edition)*, *34*(1), 28-47. https://doi.org/10.1177/2051570719829432

Cheshin, A., Amit, A., & Van Kleef, G. A. (2018). The interpersonal effects of emotion intensity in customer service: Perceived appropriateness and authenticity of attendants' emotional displays shape customer trust and satisfaction. *Organizational Behavior and Human Decision Processes*, *144*, 97-111.

Chin, H., & Yi, M. Y. (2022). Voices that care differently: Understanding the effectiveness of a conversational agent with an alternative empathy orientation and emotional expressivity in mitigating verbal abuse. *International Journal of Human–Computer Interaction*, *38*(12), 1153-1167.

Chin, W. W. (1998). The partial least squares approach to structural equation modeling. *Modern Methods for Business Research*, *295*(2), 295-336.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Coker, K. K., & Thakur, R. (2024). Alexa, may I adopt you? The role of voice assistant empathy and user-perceived risk in customer service delivery. *Journal of Services Marketing*, *38*(3), 301-311.

Cowan, B. R., Branigan, H. P., Obregón, M., Bugis, E., & Beale, R. (2015). Voice anthropomorphism, interlocutor modelling and alignment effects on syntactic choices in human- computer dialogue. *International Journal of Human-Computer Studies*, *83*,

27-42.

Csikszentmihalyi, M. (1997). Finding flow: The psychology of engagement with everyday life. *Basic Book*.

Cuciniello, M., Amorese, T., Greco, C., Callejas Carrión, Z., Vogel, C., Cordasco, G., & Esposito, A. (2023). A Synthetic Voice for an Assistive Conversational Agent: A Survey to Discover Italian Preferences regarding Synthetic Voice's Gender and Quality Level. *Human Behavior and Emerging Technologies*, *2023*(1), 8858268.

Cuddy, A. J., Glick, P., & Beninger, A. (2011). The dynamics of warmth and competence judgments, and their outcomes in organizations. *Research in organizational behavior*, *31*, 73-98.

Efthymiou, F., Hildebrand, C., de Bellis, E., & Hampton, W. H. (2024). The power of AI-generated voices: How digital vocal tract length shapes product congruency and ad performance. *Journal of Interactive Marketing*, *59*(2), 117-134.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3-4), 169-200. https://doi.org/10.1080/02699939208411068

Fakhimi, A., Garry, T., & Biggemann, S. (2023). The Effects of Anthropomorphised Virtual Conversational Assistants on Consumer Engagement and Trust During Service Encounters. *Australasian Marketing Journal*, *31*, 314-324. https://doi.org/10.1177/14413582231181140

Fornell, C., & Larcker, D. F. (1981). Structural Equation Models with Unobservable Variables and Measurement Error: Algebra and Statistics. *Journal of Marketing Research*, *18*(3), 382-388. https://doi.org/10.2307/3150980

Fuller, R. M., Harding, M. K., Luna, L., & Summers, J. D. (2022). The impact of E-commerce capabilities on online retailer performance: Examining the role of timing of adoption. *Information & Management*, *59*(2), 103584.

García-Madariaga, J., Moya, I., Recuero, N., & Blasco, M.-F. (2020). Revealing unconscious consumer reactions to advertisements that include visual metaphors. A neurophysiological experiment. *Frontiers in Psychology*, *11*, 760.

Gefen, D., & Straub, D. W. (2004). Consumer trust in B2C e-Commerce and the importance of social presence: Experiments in e-Products and e-Services. *Omega*, *32*(6), 407-424. https://doi.org/10.1016/j.omega.2004.01.006

Gelbrich, K., Hagel, J., & Orsingher, C. (2021). Emotional support from a digital assistant in technology-mediated services: Effects on customer satisfaction and behavioral persistence. *International Journal of Research in Marketing*, *38*(1), 176-193.

Ghaleb, E., Popa, M., & Asteriadis, S. (2019). Metric learning-based multimodal audio-visual emotion recognition. *IEEE Multimedia*, *27*(1), 37-48. https://doi.org/10.1109/MMUL.2019.2904622

Gong, L., & Nass, C. (2007). When a talking-face computer agent is half-human and half-humanoid: Human identity and consistency preference. *Human communication research*, *33*(2), 163-193.

Google. (2023). *Google Assistant is now available on over 2.5 billion devices*. https://blog.google/products/assistant/

Hair, J. (2009). Multivariate Data Analysis. *Faculty and Research Publications*. https://digitalcommons.kennesaw.edu/facpubs/2925

Hall, M. J., Carlson, L., & Gentry, J. W. (2024). “Aha! I knew that voice sounded familiar!”: Recognizing a non-identified voice-over endorser increases ad enjoyment via moments of insight. *Journal of Business Research*, *174*, 114495.

Ham, S.-M., Lee, H.-M., Lim, J.-H., & Seo, J. (2023). A Negative Emotion Recognition System with Internet of Things-Based Multimodal Biosignal Data. *Electronics*, *12*(20), 4321.

Henseler, J., Hubona, G., & Ray, P. A. (2016). Using PLS path modeling in new technology research: Updated guidelines. *Industrial Management & Data Systems*, *116*(1), 2-20. https://doi.org/10.1108/IMDS-09-2015-0382

Herrando, C., Jiménez-Martínez, J., Martín-De Hoyos, M. J., Asakawa, K., & Yana, K. (2022). Emotional responses in online social interactions: The mediating role of flow. *Asia Pacific Journal of Marketing and Logistics*, *35*(7), 1599-1617.

Hildebrand, C., & Bergner, A. (2021). Conversational robo advisors as surrogates of trust: Onboarding experience, firm perception, and consumer financial decision making. *Journal of the Academy of Marketing Science*, *49*(4), 659-676.

Huang, B., & Sénécal, S. (2023). How should voice assistants be heard? The mitigating effect of verbal and vocal warmth in voice assistant service failure. *The Service Industries Journal*, *43*(11-12), 806-826.

Huh, J., Kim, H.-Y., & Lee, G. (2023). "Oh, happy day!" Examining the role of AI-powered voice assistants as a positive technology in the formation of brand loyalty. *Journal of Research in Interactive Marketing*, *17*(5), 794-812.

Huh, J., Whang, C., & Kim, H.-Y. (2023). Building trust with voice assistants for apparel shopping: The effects of social role and user autonomy. *Journal of Global Fashion Marketing*, *14*(1), 5-19.

Jacob, A. (2016). Speech emotion recognition based on minimal voice quality features. *2016 International Conference on Communication and Signal Processing (ICCSP)*, 0886-0890. https://doi.org/10.1109/ICCSP.2016.7754275

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological bulletin*, *129*(5), 770.

Kim, J., Lee, K., & Kim, S. (2021). Emotional influence of virtual assistant voices on user

engagement and trust. *International Journal of Human-Computer Interaction*, *37*(8), 702-717.

Krämer, N. C., Rosenthal-von der Pütten, A. M., & Hoffmann, L. (2015). Social effects of virtual and robot companions. *The handbook of the psychology of communication technology*, 137-159.

Kudiri, K., Said, A., & Nayan, M. (2012). Emotion detection using relative amplitude-based features through speech. *2012 International Conference on Computer & Information Science (ICCIS)*, 522-525. https://doi.org/10.1109/ICCISCI.2012.6297301

Laukka, P., Juslin, P., & Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cognition & Emotion*, *19*(5), 633-653.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human factors*, *46*(1), 50-80.

Lee, M. C., Scheepers, H., Lui, A. K., & Ngai, E. W. (2023). The implementation of artificial intelligence in organizations: A systematic literature review. *Information & Management*, *60*(5), 103816.

Lucia-Palacios, L., & Pérez-López, R. (2021). Effects of home voice assistants' autonomy on instrusiveness and usefulness: Direct, indirect, and moderating effects of interactivity. *Journal of Interactive Marketing*, *56*(1), 41-54.

Luo, B., Lau, R. Y., & Li, C. (2023). Emotion-regulatory chatbots for enhancing consumer servicing: An interpersonal emotion management approach. *Information & Management*, *60*(5), 103794.

Malodia, S., Ferraris, A., Sakashita, M., Dhir, A., & Gavurova, B. (2023). Can Alexa serve customers better? AI-driven voice assistant service interactions. *Journal of Services Marketing*, *37*(1), 25-39.

Mari, A., Mandelli, A., & Algesheimer, R. (2024). Empathic voice assistants: Enhancing

consumer responses in voice commerce. *Journal of Business Research*, *175*, 114566.

McLean, G., & Osei-Frimpong, K. (2019). Hey Alexa … examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in Human Behavior*, *99*, 28-37. https://doi.org/10.1016/j.chb.2019.05.009

Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational intelligence*, *29*(3), 436-465.

Mostafa, R. B., Lages, C. R., & Shaalan, A. (2024). The dark side of virtual agents: Ohhh no! *International Journal of Information Management*, *75*, 102721.

Nass, C., & Brave, S. (2005). *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. MIT Press.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of social issues*, *56*(1), 81-103.

Nunnally, J., & Bernstein, I. (1994). *Psychometric Theory 3rd edition (MacGraw-Hill, New York)*.

Oya, R., & Tanaka, A. (2023). Touch and voice have different advantages in perceiving positive and negative emotions. *i-Perception*, *14*(2), 20416695231160420.

Pagani, M., Racat, M., & Hofacker, C. F. (2019). Adding voice to the omnichannel and how that affects brand trust. *Journal of Interactive Marketing*, *48*(1), 89-105.

Paradeda, R., Alves, Á., Torres, D., & Lima, A. (2024). The Emotions and Advice in Virtual Assistants: A Dual Study on Emotion Validation and Agent Suggestions in a Gaming Scenario. *Journal on Interactive Systems*, *15*(1), 118-129.

Pastor, M. A., Ribas, D., Ortega, A., Miguel, A., & Lleida, E. (2022). Cross-corpus speech emotion recognition with HuBERT self-supervised representation. *IberSPEECH 2022*, 76-80.

Patrizi, M., Vernuccio, M., & Pastore, A. (2021). “Hey, voice assistant!” How do users

perceive you? An exploratory study. *Sinergie Italian Journal of Management*, *39*(1), Article 1. https://doi.org/10.7433/s114.2021.10

Pepino, L., Riera, P., Gauder, L., Gravano, A., & Ferrer, L. (2020). Detecting distrust towards the skills of a virtual assistant using speech. *arXiv preprint arXiv:2007.15711*.

Pfister, P., & Lehmann, C. (2021). Returns on digitisation in SMEs—A systematic literature review. *Journal of Small Business & Entrepreneurship*, *35*, 574-598.

Poushneh, A. (2021). Humanizing voice assistant: The impact of voice assistant personality on consumers' attitudes and behaviors. *Journal of Retailing and Consumer Services*, *58*, 102283. https://doi.org/10.1016/j.jretconser.2020.102283

Qu, C., Brinkman, W.-P., Ling, Y., Wiggers, P., & Heynderickx, I. (2014). Conversations with a virtual human: Synthetic emotions and human responses. *Computers in Human Behavior*, *34*, 58-68. https://doi.org/10.1016/j.chb.2014.01.033

Ramesh, A., & Chawla, V. (2022). Chatbots in marketing: A literature review using morphological and co-occurrence analyses. *Journal of Interactive Marketing*, *57*(3), 472-496.

Reeves, B., & Nass, C. (1997). The Media Equation: How People Treat Computers, Television, ? New Media Like Real People ? Places. *Computers and Mathematics with Applications*, *5*(33), 128.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161-1178. https://doi.org/10.1037/h0077714

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*(5), 805-819. https://doi.org/10.1037/0022-3514.76.5.805

Sandoval, P. S., & García-Madariaga, J. (2024). Impact of emotional appeal on non-profit advertising: A neurophysiological analysis. *Journal of Consumer Behaviour*, *23*(1),

203-217.

Scherer, K. R. (2003). *Introduction: Cognitive components of emotion.* Oxford University Press.

Schroeder, J., & Epley, N. (2015). The sound of intellect: Speech reveals a thoughtful mind, increasing a job candidate's appeal. *Psychological science*, *26*(6), 877-891.

Sezgin, M. C., Gunsel, B., & Kurt, G. K. (2012). Perceptual audio features for emotion detection. *EURASIP Journal on Audio, Speech, and Music Processing*, *2012*, 1-21.

Shin, J.-G., Choi, G.-Y., Hwang, H.-J., & Kim, S.-H. (2021). Evaluation of Emotional Satisfaction Using Questionnaires in Voice-Based Human–AI Interaction. *Applied Sciences*, *11*(4), 1920.

Song, X., Xu, B., & Zhao, Z. (2022). Can people experience romantic love for artificial intelligence? An empirical study of intelligent assistants. *Information & Management*, *59*(2), 103595.

Statista Staff. (2023). *Global voice recognition market 2029*. Statista. https://www.statista.com/statistics/1133875/global-voice-recognition-market-size/

Torre, I., Goslin, J., & White, L. (2020). If your device could smile: People trust happy-sounding artificial agents more. *Computers in Human Behavior*, *105*, 106215.

Van der Heijden, H. (2004). User acceptance of hedonic information systems. *MIS quarterly*, 695-704.

Voicebot.ai. (2023). *Voice Assistant Demographics: Who Uses Voice Assistants?* https://voicebot.ai/voice-assistant-industry-market-statistics/

Wu, R., Liu, J., Chen, S., & Tong, X. (2023). The effect of E-commerce virtual live streamer socialness on consumers' experiential value: An empirical study based on Chinese E-commerce live streaming studios. *Journal of Research in Interactive Marketing*, *17*(5), 714-733.

Xiong, Y., Shi, Y., Pu, Q., & Liu, N. (2024). More trust or more risk? User acceptance of artificial intelligence virtual assistant. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *34*(3), 190-205.

Zhang, S., Meng, Z., Chen, B., Yang, X., & Zhao, X. (2021). Motivation, social emotion, and the acceptance of artificial intelligence virtual assistants—Trust-based mediating effects. *Frontiers in Psychology*, *12*, 728495.

Zhang, Y., Cao, Y., Proctor, R. W., & Liu, Y. (2023). Emotional experiences of service robots' anthropomorphic appearance: A multimodal measurement method. *Ergonomics*, *66*(12), 2039-2057.

3.7. Appendix

Table 3.8. Scales

Item	Scales
Positive Emotions - Adapted from Ekman (1992)	
Hearing the voice of this assistant, I felt…	
PE1	Empathy
PE2	Affection
PE3	Fondness
PE4	Attraction
Negative Emotions - Adapted from Ekman (1992)	
Hearing the voice of this assistant, I felt…	
NE1	Aversion
NE2	Disgust
NE3	Frustation
NE4	Sadness
Emotional Response - Adapted from Bradley & Lang (1994) and Kim et al. (2021)	
The assistant’s voice seemed to me…	
ER1	Very negative (1) - Very positive (7)
This assistant's voice has made me feel…	
ER2	Very relaxed (1) - Very nervous (7)
Human Like Voice - Adapted from Patrizi et al. (2021)	
HLV1	The assistant's voice sounds human
HLV2	The quality of his voice seems human
HLV3	The tone of this VA's voice seems human
Trust - Adapted from Charalambous et al. (2016)	
TRU1	The assistant's voice makes me feel comfortable
TRU2	The assistant's voice would make me trust them
TRU3	I would feel comfortable interacting with this VA

STUDY 4

The Impact of Voice Type on Purchase Intention in Voice Shopping

4.1. Introduction

Voice assistants (VAs) have rapidly transformed the way users interact with digital technology, becoming an essential part of everyday life. Recent statistics indicate that by the end of 2024, over 8.4 billion digital voice assistants will be in use worldwide, surpassing the global population (Statista, 2023). VAs, powered by artificial intelligence (AI), natural language processing (NLP), and machine learning, have extended their functionalities beyond simple tasks such as weather updates and reminders to more complex interactions, including shopping, product recommendations, and customer service (Huh & Kim, 2022) among their various applications, voice shopping has emerged as a transformative force in e-commerce, offering consumers a seamless and intuitive shopping experience through hands-free, conversational interfaces (Sun et al., 2019). As voice shopping continues to gain traction, it is essential to understand not only what drives consumer purchase decisions but also the social and psychological mechanisms that shape user engagement with VAs.

Despite the clear advantages of voice shopping—such as convenience, accessibility, and personalization—its broader adoption remains limited due to persistent concerns around privacy, trust, and the perceived lack of social presence in AI-driven interactions (Schultz & Paetz, 2023). Many users are hesitant to share sensitive information with voice assistants, citing fears of data breaches or misuse, which in turn undermines their sense of safety and willingness to complete purchases (Huh & Kim, 2022). These concerns are supported by neuroscience research, which shows that digital trust cues—like brand logos or security symbols—activate brain regions involved in risk assessment and decision-making, ultimately influencing purchase behavior (Casado-Aranda et al., 2019). This highlights the need for voice assistants to convey credibility and emotional security, particularly through the design of their auditory and interactive features.

Furthermore, a lack of social connection and engagement with voice assistants can make interactions feel impersonal, reducing user trust and weakening purchase intention (PI) (Poushneh, 2021). Since trust is a fundamental driver of consumer adoption, users must feel secure, emotionally connected, and confident in the assistant's ability to meet their needs.

Since voice is the primary channel through which users assess and form perceptions of virtual assistants, understanding how specific voice attributes influence trust, perceived

safety, emotional engagement, and social attraction is essential for explaining their impact on consumer decision-making. In the absence of these psychological and social factors, purchase intention tends to weaken, and the likelihood of users engaging in word-of-mouth (WOM) behavior declines—ultimately limiting the reach and effectiveness of voice shopping in digital commerce. WOM has long been recognized as a powerful driver of consumer decisions, with peer recommendations shaping product evaluations and purchase intentions (Berger, 2013; Hennig-Thurau et al., 2004). It is therefore important to examine whether voice assistants can moderate or enhance these effects, acting as intermediaries between recommendations and consumer action. Recent studies confirm that WOM from credible online opinion leaders strongly influences purchasing decisions, especially when users perceive the source as trustworthy and knowledgeable (Tobon & García-Madariaga, 2021). This suggests that voice assistants, when designed effectively, may serve as persuasive agents in the delivery of product recommendations.

Prior research on voice assistants has predominantly focused on their usability, convenience, and personalization in digital commerce, emphasizing aspects such as perceived ease of use, perceived usefulness, and data security concerns (Sun et al., 2019). Studies have found that VAs enhance customer satisfaction by providing efficient, hands-free solutions for shopping and offering personalized recommendations based on previous behaviors (Keerthi, 2024). These factors have positioned VAs as valuable tools for improving the overall digital shopping experience.

Despite the growing adoption of voice commerce, there remains limited understanding of how psychological mechanisms such as trust, emotional engagement, and social attraction shape user interactions with VAs and, ultimately, influence PI. More specifically, the role of voice characteristics in shaping these psychological perceptions has not been fully explored, leaving an important gap in determining to what extent the type of voice contributes to consumer engagement, purchase decision, and WOM behaviors.

The influence of voice type on key variables such as trust and emotional responses (ER) remains an underexplored area, despite their critical role in shaping consumer behavior. Emotional responses, defined as the combination of valence (positive/negative) and arousal (intensity) in reaction to stimuli, play a crucial role in shaping user engagement and decision-making processes. Despite evidence that emotions significantly impact human-computer interactions (Rzepka et al., 2023), few studies have directly linked positive

and negative emotional responses to purchase behavior in voice shopping contexts (Maduku et al., 2024; Zhu et al., 2022). Furthermore, there is little insight into how voice type interacts with other psychological variables, such as trust and social attraction, to mediate purchase decisions. Understanding these complex mediation and moderation effects is essential to advancing knowledge on how voice assistants influence purchase intention and WOM effects in digital commerce.

Research has yet to integrate the simultaneous influence of human-like voice, trust, safety, and social attraction—all of which play a crucial role in fostering engagement with VAs. While trust is widely acknowledged as essential for AI adoption, there is limited knowledge on how auditory elements contribute to trust formation and, consequently, affect PI (Schultz & Paetz, 2023). Similarly, social attraction—the extent to which users find interaction with a VA appealing and enjoyable—has been largely overlooked in voice commerce research. However, since voice characteristics directly shape user perceptions of trust, safety, and social attraction, it is essential to examine how different types of voice influence these factors and, in turn, impact PI. Prior research has shown that emotionally engaging environments can enhance user involvement and drive online purchase decisions (Tobon et al., 2020).

The type of voice used by a virtual assistant is a critical factor shaping user perceptions and behavioral responses in voice shopping. Voice characteristics can significantly influence how users perceive trust, safety, emotional connection, and social attraction (Bawack et al., 2021). As the human voice inherently conveys emotional and social cues, it has the power to evoke empathy and engagement—or, alternatively, aversion and discomfort. Neuropsychological research supports this emotional dimension, showing that secure digital environments activate brain areas associated with trust and reward, while risky settings trigger responses related to anxiety and conflict, underscoring the deep emotional relevance of perceived safety (Casado-Aranda et al., 2018). Furthermore, studies indicate that voice qualities play a central role in determining whether users perceive a virtual assistant as warm, trustworthy, and human-like—or as cold and mechanical—ultimately affecting user engagement and interaction quality (Tschopp & Sassenberg, 2024). Yet despite these insights, the specific ways in which voice characteristics shape these psychological perceptions and translate into purchase intention remain underexplored.

This chapter aims to analyze the factors that influence PI in the context of VAs, both

directly and indirectly, with voice type as the central reference point. Given that the assistant's voice fundamentally shapes the user's relationship with the VA, it serves as the starting point of our analysis. Rather than focusing solely on usability and functional benefits, this research investigates how voice type plays a key role in shaping purchase intentions, particularly through psychological and emotional factors—such as perceived human-likeness, emotional responses, perceived safety, trust, and social attraction. These variables are analyzed not only as key components of user-VA interactions but also as mediators that explain how voice type influences purchase decisions.

Additionally, we examine how behavioral intention—users' pre-existing attitudes toward VAs—impacts final PI, with voice type acting as a moderator in this process. Prior research has shown that behavioral intention significantly influences technology adoption and purchase behaviors, as users' initial attitudes shape their willingness to engage with AI-driven commerce (Venkatesh et al., 2012; Bawack et al., 2021).

While prior research has extensively studied AI-human interactions in digital commerce, the role of voice type in shaping consumer influence mechanisms remains underexplored. Given that WOM is a widely recognized driver of consumer behavior, we further explore how the type of voice moderates the relationship between users' initial perceptions of VAs and their likelihood of engaging in WOM, and how this, in turn, affects PI.

By addressing these gaps, this chapter contributes to both theoretical advancements in voice commerce research and practical insights for optimizing consumer engagement with VAs, enhancing trust, emotional connection, and long-term adoption in digital shopping experiences.

The primary objective of this chapter is to explore how the type of voice used by VAs influences consumer behavior, specifically focusing on PI. Beyond examining direct effects, this research aims to identify the underlying psychological and emotional mechanisms that mediate this relationship, offering a more comprehensive understanding of how users perceive voice assistants and how these perceptions translate into purchase-related behaviors. Given the growing reliance on voice assistants in electronic commerce, it is essential to investigate how these factors shape trust, emotional engagement, and perceived usefulness, and how these perceptions, in turn, drive purchase decisions in voice commerce contexts.

This chapter builds upon the Social Response Theory (SRT), which posits that individuals respond to technological interfaces with social behaviors similar to those seen in human-to-human interactions (Strathmann et al., 2020). In the context of voice shopping, users often attribute human-like qualities to their voice assistants, forming emotional connections that influence their shopping behaviors (Poushneh, 2021). Research suggests that voice assistants exhibiting human-like traits can enhance perceived social presence, increase trust, and foster a sense of security in voice-based interactions (Huh et al., 2023; Poushneh, 2021). However, the role of voice type in shaping these perceptions remains understudied, particularly in how different voice characteristics influence social presence, trust, and security, and how these, in turn, drive PI. While these elements have been explored separately, research has yet to integrate their combined influence with voice type as a key determinant of PI and advocacy behaviors in voice shopping.

This chapter employs a multi-model mediation framework to examine the psychological pathways through which voice type shapes consumer decision. By addressing these research questions, the chapter offers a comprehensive understanding of the psychological and emotional mechanisms underlying purchase decisions in voice shopping. These insights contribute to theoretical advancements in voice commerce research and provide practical recommendations for businesses aiming to optimize voice assistants, increase trust and engagement, and enhance consumer relationships in e-commerce environments. To guide this research, the following key questions are proposed:

- How does the type of voice used by virtual assistants influence consumer behavior both directly and indirectly through perceived humanity, emotional responses, and social attraction?

  → This question explores whether voice assistants with human-like qualities enhance perceived humanity and emotional engagement, and how voice type shapes these perceptions, ultimately influencing PI.
- What role do perceived safety and trust play in mediating the relationship between voice assistant characteristics and consumer behavior?

  → This question examines whether voice type affects perceptions of safety, which in turn strengthens trust in voice assistants, and how these factors jointly contribute to PI and advocacy behaviors in voice shopping.
- Does the type of voice assistant moderate the relationship between word of mouth,

behavioral intention, and PI?

→ This question investigates whether voice characteristics amplify or weaken the impact of WOM and behavioral intention on PI, clarifying whether certain voice types make product recommendations more persuasive or influence how strongly pre-existing behavioral intentions translate into actual purchases.

To achieve these objectives, the chapter employs a quantitative, experimental research design. Participants will be exposed to pre-recorded, standardized interactions with the three voice assistants (Alexa, Google Assistant, and Siri). This approach ensures uniform exposure and allows for a controlled assessment of user perceptions, emotional responses, and behavioral outcomes.

Data collection will be conducted through structured surveys capturing participants' evaluations of key psychological constructs, including perceived humanity in the voice, trust, safety, emotional responses, social attraction, behavioral intention (pre-exposure), and purchase-related behaviors (post-exposure).

The data analysis will employ a mediation analysis framework across six distinct models, each assessing direct and indirect effects of voice type on PI. The findings of this chapter are expected to provide significant theoretical contributions by extending the application of SRT in the context of voice shopping. By investigating how the type of voice shapes user perceptions—including trust, safety, social attraction, and emotional responses—this research provides a more nuanced perspective on how voice characteristics trigger human-like interactions that ultimately influence consumer decision-making in AI-driven shopping environments.

Beyond theoretical contributions, this research offers practical implications for businesses, voice assistant developers, and marketers. The insights gained will aid in designing voice assistants with specific voice characteristics that align with consumer expectations and emotional needs, fostering trust and engagement. Understanding how different types of voices—such as warm or authoritative tones—enhance perceived humanity can help developers create more effective assistants that drive PI and encourage word of mouth.

The chapter’s findings will also inform marketing strategies, enabling brands to personalize voice assistant interactions to better resonate with their target audiences. This

personalization can enhance customer satisfaction, encourage repeat purchases, and generate positive WOM, which are key factors for long-term success in the highly competitive digital commerce landscape. By optimizing the selection and design of voice types to enhance emotional and psychological engagement, businesses can strengthen consumer trust and create more compelling shopping experiences.

Additionally, this chapter provides valuable insights for electronic commerce platform optimization, offering guidance on how the type of voice used by assistants can enhance consumer trust and influence purchasing decisions. The results can help businesses address trust and privacy concerns, providing input for regulatory efforts aimed at enhancing transparency and user confidence in voice-based transactions. Given the increasing emphasis on AI ethics and responsible AI development, these insights can inform best practices for designing voice interfaces that not only prioritize functionality but also ensure that voice characteristics align with user expectations and emotional engagement.

Finally, this research contributes to the future of voice commerce by highlighting how the specific characteristics of AI-driven voices can strengthen consumer relationships. As voice assistants continue to evolve, these insights will serve as a foundation for improving next-generation voice-based interactions. By bridging the gap between academic theory and practical application, this chapter supports the continued growth and optimization of voice commerce, ensuring that the selection and design of voice types become a strategic tool for enhancing customer relationships and driving commercial success.

4.2. Type of Voice in Voice Shopping: Theoretical Framework and Hypothesis Development

4.2.1. Direct effect

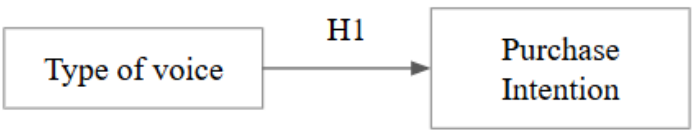
The integration of VAs in digital commerce has significantly altered consumer behavior, providing hands-free, conversational shopping experiences that enhance efficiency, personalization, and accessibility (Sun et al., 2019). As voice shopping continues to gain adoption, research has increasingly explored the factors that drive user engagement and purchase decisions. Among these, the type of voice used by VAs has emerged as a key element in shaping consumer perceptions and influencing trust, emotional responses, and overall purchase behaviors (Bawack et al., 2021).

Voice is a fundamental component of human communication, carrying social, emotional, and cognitive cues that influence how individuals interpret interactions (Schroeder et al., 2017). In the context of voice shopping, the tone, cadence, and expressiveness of a VA's voice contribute to perceptions of competence, warmth, and trustworthiness (Nass & Brave, 2005). Studies suggest that when voice assistants exhibit human-like attributes, they foster greater social engagement and perceived reliability, which are crucial for building consumer trust (Huh et al., 2023). Conversely, robotic, monotonous, or overly synthetic voices can lead to negative emotional responses and a diminished sense of connection with the assistant, ultimately affecting users' willingness to engage in purchasing behaviors (Tschopp & Sassenberg, 2024).

Prior research has highlighted the relationship between voice attributes and purchase behavior, demonstrating that voices perceived as warm, engaging, and expressive positively impact consumer trust, safety perceptions, and PI (Poushneh, 2021). Consumers are more likely to follow purchase recommendations from assistants that exhibit natural and relatable voice patterns, as these cues reduce uncertainty and enhance credibility (Schultz & Paetz, 2023). On the other hand, a mechanical or artificial-sounding voice may increase user skepticism and perceived risk, leading to lower trust and reduced willingness to make a purchase (Jang et al., 2022). Therefore, the voices of virtual assistants will have a different effect on users depending on the characteristics of the assistant's voice. Based on this, we propose H1, which addresses the direct effect of voice type on PI.

H1. The type of voice of a virtual assistant directly influences purchase intention, such that different voice types produce distinct effects on consumer purchase intention.

Figure 4.1. Model 1



4.2.2. Mediated effects

Human-Like Voice

A fundamental aspect of human-computer interaction is the extent to which users perceive technology as human-like. The SRT suggests that individuals apply social norms to

their interactions with technology, particularly when the interface exhibits human-like characteristics (Strathmann et al., 2020). In the case of voice assistants, the type of voice plays a crucial role in shaping perceptions of humanity, influencing whether users engage with the assistant as they would with a human or perceive it merely as a machine (Nass & Brave, 2005).

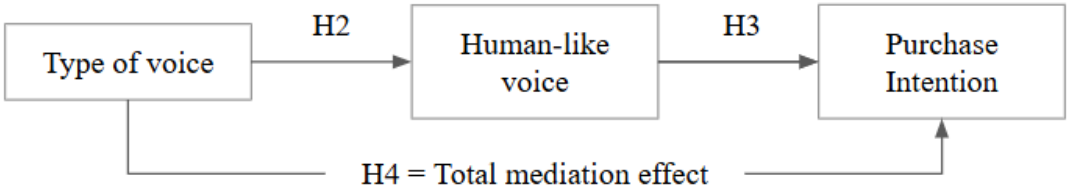
Research has shown that specific voice characteristics—such as natural intonation, warmth, emotional expressiveness, and conversational fluidity—enhance the perception of humanity in voice assistants (Schroeder et al., 2017). When a VA's voice mimics human speech patterns, users are more likely to attribute social presence and intelligence to the assistant, leading to greater engagement (Huh et al., 2023). Conversely, robotic or monotonous voices reduce perceived humanity, making the assistant feel less relatable and diminishing the overall user experience (Jang et al., 2022). Given that the perception of humanity in a voice assistant depends directly on its voice characteristics, we propose the following hypothesis:

H2. The perceived humanity of a virtual assistant depends directly on the characteristics of its voice.

Perceiving a voice assistant as human-like can have significant implications for consumer behavior. Studies indicate that higher perceived humanity in AI interfaces enhances user trust, engagement, and willingness to act upon recommendations (Poushneh, 2021). In voice shopping, users who perceive the assistant as more human-like tend to experience greater emotional connection and lower uncertainty, leading to higher PI (Tschopp & Sassenberg, 2024). Neuroimaging evidence shows that online trust signals activate brain areas related to risk and decision-making, highlighting the emotional basis of trust in digital contexts (Casado-Aranda et al., 2019). Additionally, voice assistants perceived as more human-like are more persuasive, as users are more inclined to follow suggestions from sources they perceive as socially intelligent and trustworthy (Schultz & Paetz, 2023). Given these findings, we propose the following hypothesis:

H3: The perceived humanity of a virtual assistant directly influences purchase intention, such that higher perceived humanity leads to higher purchase intention.

Figure 4.2. Model 2



H4. The relationship between the type of voice of the virtual assistant and purchase intention is mediated by the perceived humanity of the assistant's voice.

Perceived Safety and Trust

One of the main barriers to adopting voice shopping is the perception of security risks and lack of trust in voice assistants (Schultz & Paetz, 2023). Consumers are often hesitant to share personal and financial information with VAs due to concerns over data privacy, unauthorized access, and potential misuse of sensitive data (Huh & Kim, 2022). Given these concerns, perceived safety plays a key role in shaping trust, which in turn influences willingness to purchase through voice assistants.

Research suggests that the characteristics of a voice assistant's voice can influence perceived safety, as certain vocal attributes convey reliability, professionalism, and emotional warmth (Bawack et al., 2021). Human-like voices that sound natural, calm, and expressive tend to create a greater sense of security, as users feel more comfortable interacting with the assistant (Tschopp & Sassenberg, 2024). Conversely, robotic or mechanical voices may increase skepticism, as they lack the emotional and social cues that foster a sense of reassurance and reliability (Jang et al., 2022). Since users are more likely to feel safe when interacting with voice assistants that exhibit natural and engaging speech patterns, we propose the following hypothesis:

H5. The perceived safety of a virtual assistant depends directly on the characteristics of its voice.

Safety and trust are closely interconnected in human-computer interactions. Studies have shown that when users feel safe using a digital system, they are more likely to develop trust in the technology (Schroeder et al., 2017). In the case of voice assistants, perceived safety—particularly regarding privacy, data security, and system reliability—directly influences trust (Poushneh, 2021). If users feel that their personal data is protected and that

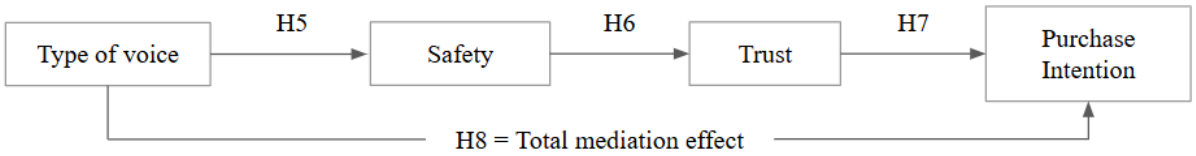
the assistant operates transparently and efficiently, they are more inclined to trust it for decision-making and purchase recommendations (Schultz & Paetz, 2023). Given the strong link between security perception and trust formation, we propose the following hypothesis:

H6. The perceived safety of a virtual assistant directly influences trust, such that higher perceived safety leads to greater trust.

Trust is a fundamental determinant of consumer adoption in digital commerce. Users who trust a voice assistant are more likely to engage in interactions, rely on its recommendations, and ultimately make purchases (Huh et al., 2023). Trust also reduces perceived risk, making consumers feel more confident about their shopping decisions in a voice commerce environment (Jang et al., 2022). On the other hand, low trust levels can create hesitation, leading to reduced engagement and lower purchase intention. Since trust is a key driver of purchase behaviors, we propose the following hypothesis:

H7. Trust in a virtual assistant directly influences purchase intention, such that higher trust leads to higher purchase intention.

Figure 4.3. Model 3



H8. The relationship between the type of voice of the virtual assistant and purchase intention is sequentially mediated by safety and trust.

Emotional Responses and Social Attraction

Emotions play a fundamental role in consumer decision-making, particularly in human-computer interactions where the design of the interface can evoke both positive and negative emotional responses (ER). Emotional responses are defined as the combination of valence (positivity or negativity) and arousal (intensity) that influences user behavior (Rzepka et al., 2023). In the context of voice assistants, ER refers to the emotional reaction users experience when interacting with the assistant's voice, which can range from comfort and trust to frustration and skepticism (Maduku et al., 2024).

Research suggests that voice characteristics, such as tone, expressiveness, and

warmth, can shape users' emotional reactions (Tschopp & Sassenberg, 2024). A voice that sounds natural, friendly, and engaging is more likely to evoke positive emotions, such as comfort, trust, and enjoyment, whereas a voice that sounds monotone, mechanical, or robotic may trigger negative emotions, such as discomfort, frustration, or lack of engagement (Schultz & Paetz, 2023). Since emotional responses are highly influenced by auditory cues, we propose the following hypothesis:

H9. Emotional responses to a virtual assistant depend directly on the characteristics of its voice.

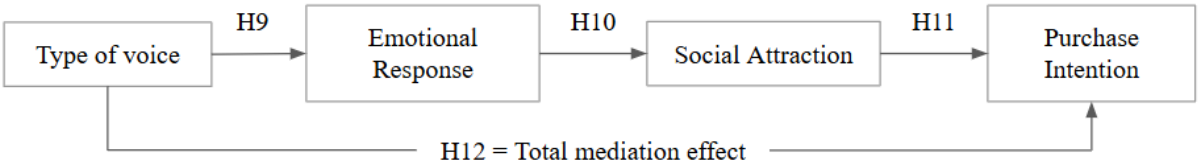
Emotional responses are closely linked to social attraction, which refers to the degree to which users find the virtual assistant engaging, likable, and appealing (Strathmann et al., 2020). When a voice assistant evokes positive emotions (e.g., warmth, trust, and enjoyment), users are more likely to perceive it as socially attractive and engaging (Poushneh, 2021). In contrast, negative emotional responses (e.g., irritation, discomfort) reduce social attraction, making the assistant feel less relatable and less enjoyable to interact with (Huh et al., 2023). Given the strong connection between emotional engagement and social attraction, we propose the following hypothesis:

H10. Emotional responses to a virtual assistant directly influence social attraction, such that more positive emotional responses lead to higher social attraction.

Social attraction plays a critical role in shaping user engagement and purchase behavior in voice shopping contexts. When users perceive a voice assistant as socially engaging and likable, they are more inclined to trust its recommendations and follow its purchase suggestions (Schroeder et al., 2017). A voice assistant that fosters a strong sense of connection and engagement can enhance consumer willingness to purchase, while a lack of social attraction may create psychological distance, reducing purchase intention (Jang et al., 2022). Since social attraction drives engagement and decision-making, we propose the following hypothesis:

H11. Social attraction to a virtual assistant directly influences purchase intention, such that higher social attraction leads to a greater likelihood of making a purchase.

Figure 4.4. Model 4



H12. The relationship between the type of voice of the virtual assistant and purchase intention is sequentially mediated by emotional responses and social attraction.

4.2.3. The Moderating Role of Voice Type

Consumer decision-making in voice shopping is shaped by trust, emotional engagement, and prior attitudes, as well as the recommendations users generate from their experience with the assistant. Two key factors influencing purchase behavior in this context are Behavioral Intention (BI)—a consumer's predisposition to make a purchase using voice assistants—and WOM—the likelihood of recommending the assistant based on user experience. Research has shown that pre-existing attitudes toward technology significantly predict adoption behaviors and purchasing decisions, particularly in AI-mediated interactions (Bawack et al., 2021; Venkatesh et al., 2012). Similarly, WOM plays a crucial role in shaping consumer trust and purchase likelihood, as positive recommendations serve as persuasive social proof (Berger, 2014; Hennig-Thurau et al., 2004). However, the extent to which these factors translate into actual purchase intention depends on the assistant's voice characteristics, which influence the persuasiveness of pre-existing purchase intentions and users' willingness to recommend the assistant (Schroeder et al., 2017).

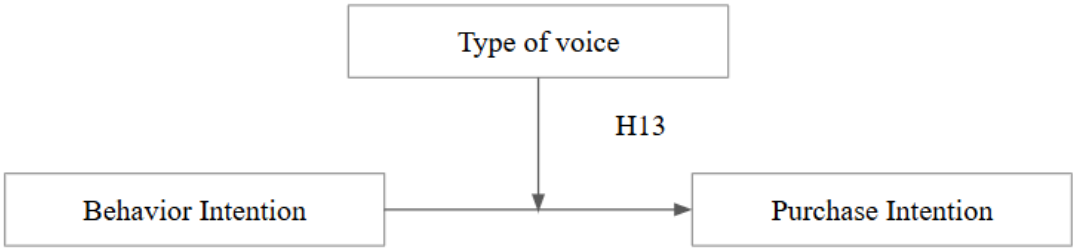
Behavioral Intention is a key predictor of purchase intention, reflecting a consumer's readiness to buy based on attitudes, motivations, and past experiences (Bawack et al., 2021). While strong behavioral intentions typically lead to higher purchase likelihood, research suggests that contextual factors and message delivery styles can amplify or weaken this effect (Schroeder et al., 2017). In voice shopping, the assistant's voice characteristics determine how effectively behavioral intention translates into a final purchase decision, as voice expressiveness enhances persuasiveness, engagement, and perceived credibility (Pias et al., 2024).

Certain voice characteristics may reinforce a user's pre-existing behavioral intention,

increasing confidence in the purchase decision (Tschopp & Sassenberg, 2024). Conversely, other voice types may introduce psychological distance, making users less certain about their decision and reducing purchase intention (Jang et al., 2022). Since the strength of the BI–purchase intention relationship varies depending on the assistant's voice characteristics, we propose the following hypothesis:

H13. The direct and positive relationship between behavioral intention and purchase intention is moderated by the type of voice of the assistant.

Figure 4.5. Model 5



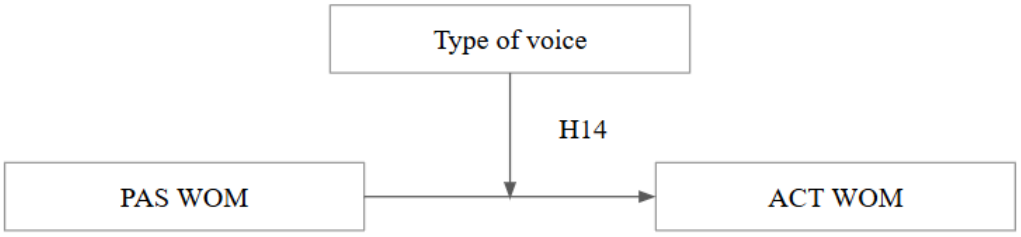
WOM is not a static construct; it evolves based on user experiences. Prior to interacting with a voice assistant, users may hold a general perception or expectation about its effectiveness, which we define as Passive WOM. This refers to pre-existing attitudes or opinions users have formed based on previous exposure to information, such as online reviews, brand reputation, or indirect experience (East et al., 2008; Wang & Yu, 2017).

However, once users interact with the assistant, their real-time experience shapes their likelihood of actively recommending or discouraging its use, which we define as Active WOM. Unlike Passive WOM, which is shaped by external information sources, Active WOM emerges from direct interaction and firsthand evaluation of the assistant's performance.

The degree to which Passive WOM translates into Active WOM depends on the characteristics of the assistant's voice. A natural and engaging voice may reinforce or amplify initial expectations, ensuring consistency between Passive and Active WOM (Kim et al., 2021). Conversely, a robotic or mechanical-sounding voice may disrupt this continuity, leading users to reassess their initial opinions and alter their Active WOM. This aligns with theories of cognitive fluency and expectation violation, which suggest that discrepancies between pre-existing beliefs and actual experience lead to reevaluations in attitude and behavior (Kahneman, 2011; Reber et al., 2004). Thus, we propose the following hypothesis:

H14. The direct and positive relationship between Passive WOM (pre-stimulus WOM) and Active WOM (post-stimulus WOM) is moderated by the type of voice of the assistant.

Figure 4.6. Model 6

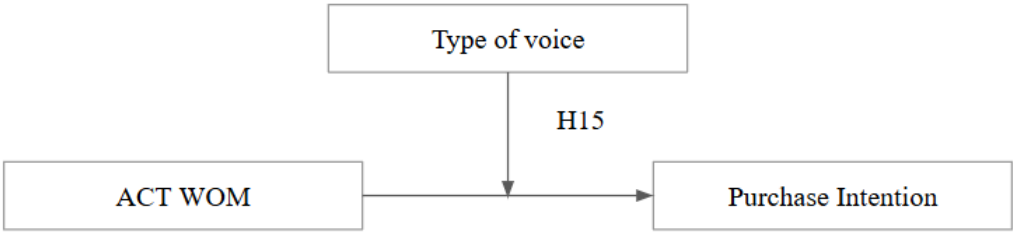


Once users have formed an Active WOM assessment after interacting with the assistant, their likelihood of recommending it can directly impact purchase intention. Unlike Passive WOM, which is shaped by prior beliefs, Active WOM is based on the user's direct evaluation of the assistant's performance, influencing how much they trust its recommendations (Schultz & Paetz, 2023).

However, the strength of the relationship between Active WOM and purchase intention is not universal. Voice characteristics may reinforce the credibility of Active WOM, making users more receptive to recommendations and increasing their purchase intention (Huh et al., 2023). In contrast, a voice that lacks engagement may weaken this relationship, leading users to view recommendations with skepticism or detachment (Jang et al., 2022). Thus, we propose the following hypothesis:

H15. The direct and positive relationship between Active WOM (post-stimulus WOM) and purchase intention is moderated by the type of voice of the assistant.

Figure 4.7. Model 7



4.3. Research Methodology

As detailed in Chapter 3, this chapter employs an experimental design to examine how the type of voice in virtual assistants influences key psychological constructs, including human-like voice perception, emotional responses, perceived safety, trust, and social attraction, and how these factors ultimately impact purchase intention and word of mouth.

Participants were exposed to synthetic voices from three widely used virtual assistants—Siri, Google Assistant, and Alexa—to assess their perceptions and behavioral responses. A six-condition experimental design was implemented, with participants randomly assigned to listen to either positive or negative messages narrated by one of the assistants. Control measures were included to validate the voice type manipulation and mitigate potential biases arising from prior recognition of virtual assistants.

The sample consisted of 900 participants with an equal gender distribution and an age range of 18 to 50 years, selected based on their high level of interaction with virtual assistants. Data collection took place in June 2022 through an online platform provided by a market research firm.

For more detailed information on the experimental design, participant selection, control variables, and statistical analysis, refer to Chapter 3.

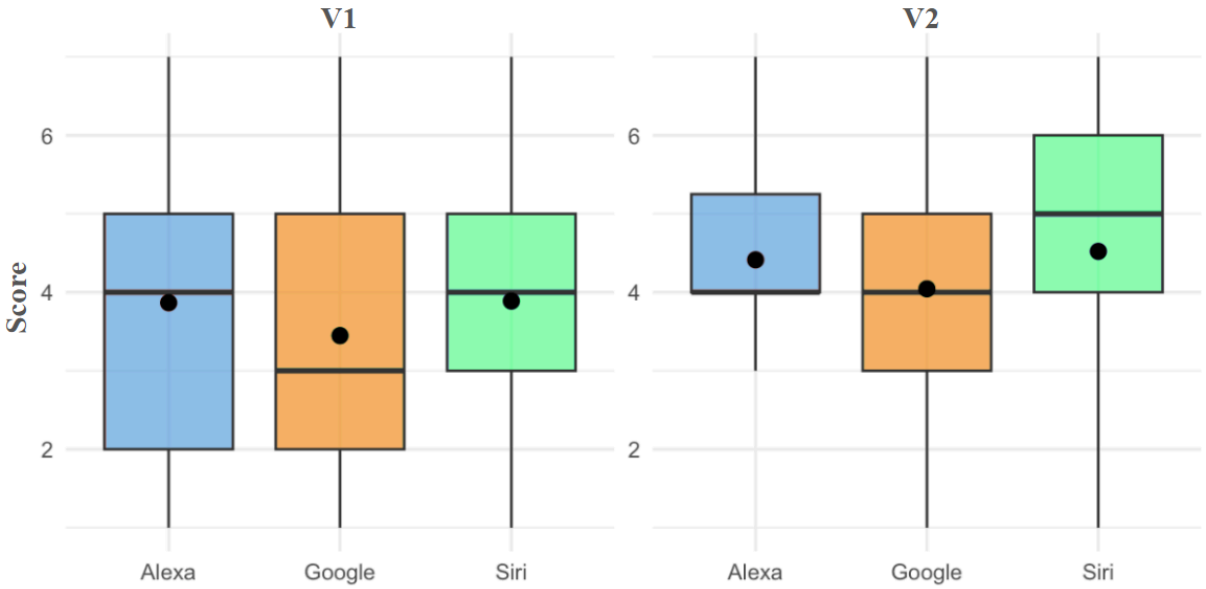
4.3.1. Voice characteristics analysis

To assess voice characteristics, we used two variables based on the scale from Bartneck et al. (2009); a semantic differential scale measuring voice from robotic to human (V1) and another measuring voice from irritating to friendly (V2). Research has consistently shown that voice characteristics significantly shape user perceptions of AI-based voice assistants. More natural and expressive voices foster greater trust and engagement, while robotic or monotonous voices reduce credibility and emotional connection (Pias et al., 2024). These characteristics influence consumer decision-making, as perceived human-likeness and trust impact purchase intention and overall user satisfaction (Huh et al., 2023). Similarly, tone and vocal characteristics affect whether users perceive a voice assistant as a reliable and engaging shopping companion (Grewal et al., 2022).

Figure 4.8 compares the three voice assistants based on two perceptual dimensions. In

terms of robotic to human (V1), Siri (M = 3.89, SD = 1.81) and Alexa (M = 3.87, SD = 1.79) receive similar ratings, while Google Assistant (M = 3.45, SD = 1.89) is perceived as significantly more robotic. This supports previous research showing that differences in speech synthesis influence how natural a voice sounds (Huh et al., 2023). For irritation vs. friendliness (V2), Siri is rated as the most pleasant (M = 4.52), significantly higher than Google Assistant (M = 4.04), while Alexa falls in between (M = 4.41). These findings highlight that although both Siri and Alexa achieve relatively high human-likeness scores, Siri maintains a distinct advantage in fostering a more pleasant and engaging auditory experience, which may contribute to its stronger performance across other behavioral measures in subsequent models.

Figure 4.8. Voice characteristics



4.3.2. Methodological approach for moderation models

To examine how voice assistants moderate key relationships in consumer decision-making, Models 5, 6, and 7 follow a consistent methodological approach, employing regression analyses with interaction terms, post hoc comparisons, and trend estimations. These models specifically investigate (1) how voice type moderates the relationship between behavioral intention and purchase intention (Model 5), (2) how voice type moderates the transition from passive WOM to active WOM (Model 6), and (3) how voice type moderates the impact of active WOM on purchase intention (Model 7). By using the same analytical framework for these models, this section provides a unified explanation of the

methodological choices, avoiding redundancy in later sections.

All statistical analyses were conducted using R (version 4.4.3), leveraging advanced regression modeling techniques to assess moderation effects across different voice assistants. A key component of this approach was the inclusion of interaction terms to determine whether the relationship between independent and dependent variables varied depending on the assistant. The following functions were used for statistical estimation and hypothesis testing:

- *lm()* – to estimate linear regression models and assess main effects and interaction terms.
- *emtrends()* – to compute marginal slopes for each voice assistant, allowing for a clearer interpretation of moderating effects.
- *pairs()* – to perform post hoc pairwise comparisons, examining whether the differences between Siri, Google Assistant, and Alexa were statistically significant.
- *interact_plot()* – to generate interaction effect visualizations, illustrating how voice type influences behavioral relationships.

Each model employed a regression framework with interaction terms, where the independent variable was tested in relation to the dependent variable (purchase intention or active WOM), with voice type included as a moderator. The statistical analysis followed a hierarchical model testing approach to systematically evaluate moderating effects:

- Baseline Model: A simple regression was estimated to test the main effect of the independent variable on the dependent variable.
- Moderation Model: An interaction term (Independent Variable × Voice Assistant) was introduced to assess whether voice type significantly moderates the relationship.
- Post Hoc Comparisons: Pairwise post hoc tests were conducted to identify significant differences between voice assistants in moderating the effects.
- Interaction Visualization: Graphical interaction plots were generated to illustrate how the relationship between independent and dependent variables varies by VA type.

The methodological approach in this study aligns with best practices in moderation analysis, incorporating interaction terms in regression models to examine how voice assistants influence consumer decision-making relationships (Hayes, 2017). Post hoc comparisons with Bonferroni corrections ensure statistical robustness by mitigating Type I

errors, while bootstrapping with 5,000 resamples enhances the reliability of interaction effects and confidence intervals, reducing sensitivity to sampling variability (Baranger et al., 2023; Preacher & Hayes, 2008).

4.3.3. Measuring variables

This chapter assessed key psychological and behavioral constructs related to VA perception, emotional responses, trust, and consumer behavior. All variables were measured using 7-point Likert scales (1 = Strongly Disagree, 7 = Strongly Agree), adapted from validated scales to align with prior research in voice shopping and consumer decision-making.

Emotional responses were evaluated through valence and arousal, capturing positivity/negativity and intensity of user reactions. Valence was measured with "The voice of this assistant seemed to me…" (1 = Very negative to 7 = Very positive), based on HCI standards (Russell, 1980). Arousal was assessed with "The voice of this assistant made me feel…" (1 = Very relaxed to 7 = Very nervous), reflecting emotional engagement and its impact on trust (Bradley & Lang, 1994).

Perceived human-likeness was measured using Patrizi et al. (2021) Human-Like Voice scale, evaluating intonation, expressiveness, and conversational flow. Trust in the assistant was assessed with an adapted Trust in AI Systems scale (Charalambous et al., 2016), capturing reliability, competence, and transparency in AI-driven interactions. Participants were instructed to focus on voice characteristics rather than system functionality. Social Attraction was measured using McLean & Osei-Frimpong (2019) scale, assessing likability, warmth, and relational connection with the VA. Perceived Safety was evaluated using the same authors' validated scale, measuring data security confidence, privacy protection, and safety perceptions in voice-assisted shopping.

Behavioral Intention, capturing readiness to engage in voice shopping, was measured using Bawack et al. (2021) scale. Purchase Intention, assessing confidence in AI-driven transactions and willingness to follow VA-generated recommendations, was also adapted from Bawack et al. (2021). Active WOM, the likelihood of recommending products discovered through the VA, was measured using Moriuchi (2019) scale. Passive WOM, users' pre-existing opinions before interaction, was assessed with Hedeman (2021) scale,

allowing comparison between initial perceptions and post-interaction WOM behaviors.

All measurement scales used in this chapter were adapted from prior research and validated through exploratory factor analysis (EFA) to ensure reliability and construct validity. The detailed measurement scales and their corresponding items can be found in Appendix (Table 4.3).

4.4. Analyses and results

To examine the proposed hypotheses, we analyzed the collected data using mediation models, which primarily involve regression analyses to explore the relationships between variables and assess the indirect effects. Mediation models allow for the estimation of how an independent variable influences a dependent variable through one or more mediators. To ensure the dimensionality, reliability, and validity of the scales, we conducted an exploratory factor and a confirmatory factor analysis using the statistical software SPSS (version 24) and Smart PLS 4. For internal reliability, we measured Cronbach's alpha of all variables, exceeding the required minimum value of 0.70, indicating that our variables had satisfactory levels of reliability (Henseler et al., 2016). For convergent validity, item loadings were checked, exceeding the ideal value of 0.70 (Hair, 2009). The composite reliability values of all variables were greater than 0.70, indicating appropriate internal consistency of the variables (Aguirre-Urreta et al., 2013), while the average variance extracted (AVE) showed consistency with values higher or equal to 0.50, indicating that each of the variables captured enough variance over the amount of variance from measurement errors (Fornell & Larcker, 1981). We can conclude that the scales met the requirements of reliability. The results of these analyses are shown in Table 4.1.

Table 4.1. Measurement model results

Variable	Item	Mean	SD	λ	t-value	CA	CR	AVE
ER	ER1	6.254	0.698	0.888	101.848***	0.745	0.746	0.797
	ER2	6.233	0.660	0.898	123.619***			
PUR	PUR1	6.254	0.698	0.888	101.848***	0.745	0.746	0.797
	PUR2	6.233	0.660	0.898	123.619***			
HLV	HLV1	5.112	1.035	0.922	120.872***	0.924	0.925	0.869
	HLV2	5.159	1.067	0.943	191.976***			
	HLV3	5.267	1.006	0.932	188.247***			
TRU	TRU1	5.409	0.876	0.921	149.322***	0.905	0.907	0.841
	TRU2	5.380	0.934	0.928	162.385***			
	TRU3	5.510	0.872	0.902	96.086***			
BE	BE1	5.09	1.602	0.833	33.199***	0.729	0.724	0.650
	BE2	5.18	1.605	0.850	40.420***			
	BE3	3.35	1.756	0.730	25.019***			
SAF	SAF1	4.427	1.541	0.869	96.535***	0.732	0.868	0.767
	SAF2	3.151	1.863	0.882	99.032***			
AWOM	AWOM1	3.77	1.81	0.937	160.602***	0.923	0.924	0.867
	AWOM2	3.65	1.82	0.941	204.086***			
SATR	SATR1	3.303	1.913	0.938	190.586***	0.856	0.857	0.874
	SATR2	4.017	1.814	0.931	157.589***			

Notes: CA: Cronbach's alpha; CR: composite reliability; AVE: average variance extracted.
***sig. < 0.01.

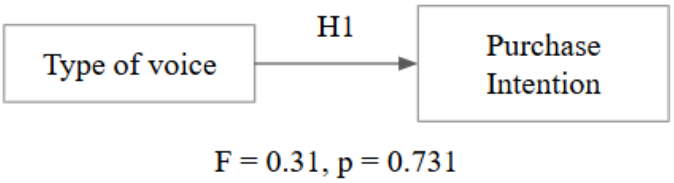
ER: Emotional Responses; PUR: Purchase Intention; HLV: Human-Like Voice; TRU: Trust; BE: Behavior Intention; SAF: Safety; AWOM: Active WOM; SATR: Social Attraction

4.4.1. Model 1

To examine whether voice type directly influences purchase intention, a one-way ANOVA was conducted, comparing purchase intention scores across different voice conditions. The results indicated that the effect of voice type on purchase intention was not statistically significant ($F = 0.31$, $p = 0.731$).

These findings suggest that differences in voice type alone do not produce significant variations in consumers' purchase intentions so H1 is not supported. While prior research has highlighted the importance of voice characteristics in shaping trust and engagement, the lack of a direct effect in this chapter implies that other psychological mechanisms, such as perceived humanity, trust, or emotional responses, may mediate this relationship.

Figure 4.9. Model 1 results



4.4.2. Model 2

Since Model 1 did not support a direct relationship between voice type and purchase intention, we conducted a bootstrapped mediation analysis (5,000 resamples) using PROCESS Model 4 (Hayes, 2017) to assess whether perceived human-likeness mediates this effect.

The first step examined whether voice type influences perceived human-likeness (H2), revealing a significant effect ($\beta = 0.42$, SE = 0.11, $p = 0.018$). The mean pratings for each assistant were: Siri (M = 4.35), Alexa (M = 4.18), and Google Assistant (M = 3.82). Assistants with voices rated as more human-like were perceived as more engaging and trustworthy, whereas those with robotic or monotone voices were rated lower. As discussed in Section 3.2, Siri ranked highest in human-likeness, followed by Alexa, while Google Assistant received the lowest ratings. These differences suggest that Siri's perceived human-like qualities enhance trust, engagement, and purchase intention, whereas Google Assistant's lower human-likeness may weaken these effects. This aligns with prior research showing that human-like voices enhance trust and decision-making (Patrizi et al., 2021).

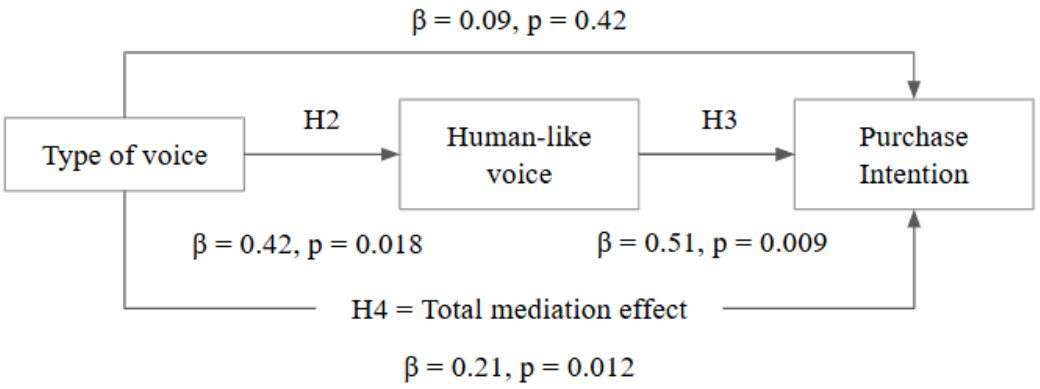
Next, we analyzed the relationship between human-likeness and purchase intention (H3), confirming a significant association ($\beta = 0.51$, SE = 0.10, $p = 0.009$). Users perceiving the assistant as more human-like were more likely to make a purchase, consistent with prior findings on AI social presence and consumer engagement (Huh et al., 2023).

The indirect effect of voice type on purchase intention through human-likeness (H4)

was significant (β = 0.21, 95% CI [0.07, 0.38], p = 0.012), while the direct effect was non-significant (β = 0.09, SE = 0.12, p = 0.42), confirming full mediation. These findings suggest that voice type shapes user perceptions, but it is the perception of human-likeness—rather than voice type itself—that drives purchase intention.

These results confirm Hypotheses H2 to H4, reinforcing the pivotal role of voice characteristics in consumer decision-making. Beyond functionality, the voice of a virtual assistant serves as a psychological driver of trust, emotional connection, and purchase behavior. Siri, as the most human-like assistant, fosters the strongest engagement and purchase likelihood, followed by Alexa, while Google Assistant's lower human-likeness ratings may reduce its persuasive impact.

Figure 4.10. Model 2 results



4.4.3. Model 3

This analysis examined whether voice type influences perceived safety (H5). The results confirmed a significant effect (β = 0.39, SE = 0.10, p = 0.015). The mean ratings for each assistant were: Siri (M = 4.28), Alexa (M = 4.12), and Google Assistant (M = 3.79).

To examine the primary mediation pathway, we examined whether safety contributes to trust formation (H6), confirming a strong association (β = 0.48, SE = 0.09, p = 0.008). Users who feel secure interacting with an assistant are more likely to trust its recommendations, strengthening its role in facilitating consumer decisions. Moreover, trust significantly predicted purchase intention (H7) (β = 0.52, SE = 0.09, p = 0.007), underscoring its key role in AI-driven commerce.

A bootstrapped mediation analysis (5,000 resamples) tested the sequential mediation

pathway (Type of Assistant → Safety → Trust → Purchase Intention), which yielded the strongest effect (β = 0.25, 95% CI [0.10, 0.37], p = 0.006), confirming that feeling safe enhances trust, which ultimately increases purchase intention. These results provide strong support for H8, indicating a full mediation process where safety fosters trust, which in turn drives purchasing behavior.

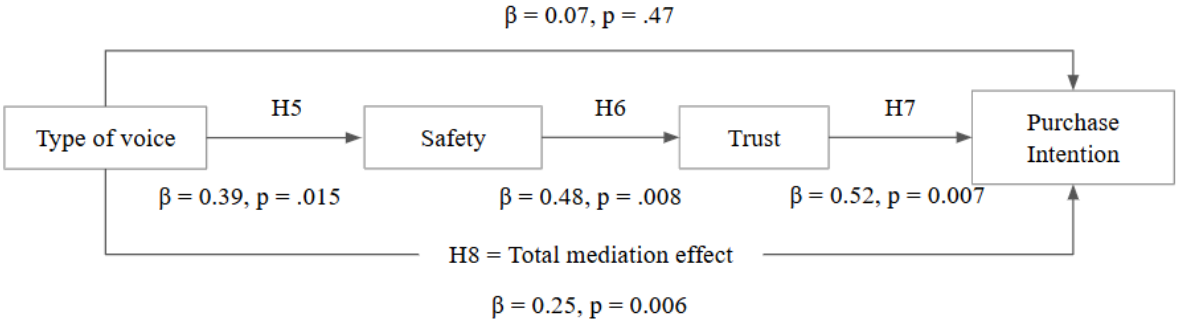
Additionally, besides the main sequential mediation pathway, we examined the separate effects of safety and trust individually. Voice type significantly influenced trust (β = 0.44, SE = 0.11, p = 0.012). The mean ratings for each assistant were: Siri (M = 4.32), Alexa (M = 4.15), and Google Assistant (M = 3.81). Siri's higher ratings likely enhances trust and credibility, while Google Assistant, with lower ratings, may struggle to build the same level of confidence (Huh et al., 2023).

- Safety as the sole mediator: The effect was significant (β = 0.14, 95% CI [0.04, 0.26], p = 0.024), suggesting that users who feel safer are more inclined to purchase.
- Trust as the only mediator: The effect was stronger (β = 0.20, 95% CI [0.08, 0.32], p = 0.010), reinforcing that trust formation is a key driver.

Notably, the direct effect of voice type on purchase intention was no longer significant (β = 0.07, SE = 0.11, p = 0.47), indicating a full mediation effect. This suggests that voice type alone does not drive purchase intention but does so through its influence on safety and trust.

These findings confirm Hypotheses H5 to H8, emphasizing the pivotal role of trust and safety in AI-driven commerce. While previous models highlighted the importance of human-likeness and emotional engagement, Model 3 reveals that trust formation is a sequential process, where perceived safety is a prerequisite for trust, and trust ultimately drives purchase behavior. Siri remains the most effective at fostering safety and trust, followed by Alexa, while Google Assistant shows the weakest influence.

Figure 4.11. Model 3 results



Indirect Effects
Effect 1: Type of Voice → Safety → Purchase Intention [0.04, 0.26]
Effect 2: Type of Voice → Trust → Purchase Intention [0.08, 0.32]
Effect 3: Type of Voice → Safety → Trust → Purchase Intention [0.10, 0.37]

4.4.4. Model 4

This analysis examined the mediating effect of perceived emotions. The results confirmed that voice type significantly influenced emotional responses (H9) (β = 0.45, SE = 0.10, p = 0.009). The mean ratings for each assistant were: Siri (M = 4.30), Alexa (M = 4.17), and Google Assistant (M = 3.85). These results suggest that assistants perceived as less robotic and less irritating, such as Siri, foster a stronger emotional connection, increasing willingness to interact and purchase. This aligns with prior findings that human-like voices enhance engagement and decision-making in AI-driven shopping (Patrizi et al., 2021). Conversely, Google Assistant's lower expressiveness and lower emotional response ratings may contribute to weaker emotional engagement and reduced purchase intention, reinforcing the importance of voice design in AI commerce.

The results also showed that emotional responses positively influence social attraction (H10) (β = 0.52, SE = 0.09, p = 0.009), suggesting that voice assistants evoking strong emotions are perceived as more socially engaging, likable, and relatable (Huh et al., 2023). Furthermore, social attraction significantly influences purchase intention (H11) (β = 0.48, SE = 0.10, p = 0.011), reinforcing that users are more inclined to trust and act on recommendations from AI assistants they perceive as socially engaging (Schroeder et al., 2017).

A bootstrapped mediation analysis (5,000 resamples) assessed the sequential mediation pathway (Type of Assistant → Emotional Response → Social Attraction →

Purchase Intention). The results confirmed the strongest indirect effect (β = 0.22, 95% CI [0.10, 0.34], p = 0.009), indicating that users who experience positive emotions when interacting with the assistant are more likely to find it socially engaging, which ultimately increases purchase intention. These findings provide strong support for H12, confirming that both emotional responses and social attraction jointly mediate the relationship between voice type and purchase behavior.

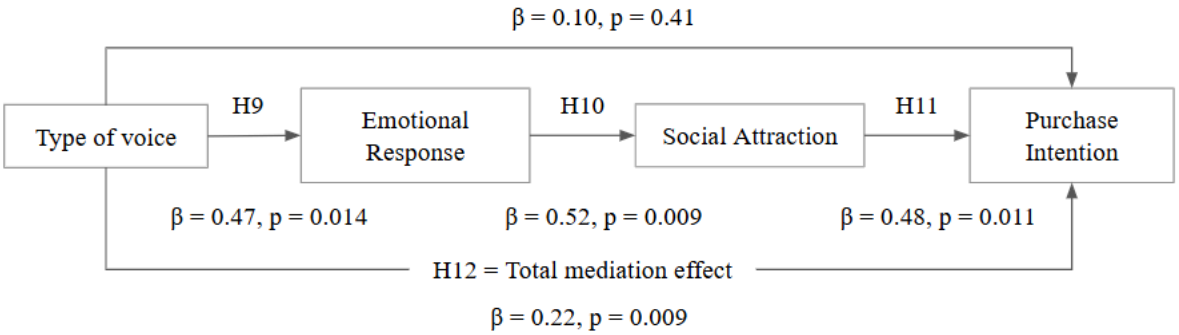
Additionally, besides the main sequential mediation pathway, we examined the separate indirect effects of emotional responses and social attraction:

- Emotional responses as the sole mediator: The effect was significant but weaker (β = 0.12, 95% CI [0.03, 0.24], p = 0.027), suggesting that while emotions play a role, they do not fully explain purchase intention.
- Social attraction as the only mediator: The effect was stronger (β = 0.15, 95% CI [0.05, 0.27], p = 0.021), reinforcing that social connection shapes purchase behavior but is not the sole driver.

Finally, the direct effect of voice type on purchase intention was no longer significant (β = 0.10, SE = 0.11, p = 0.41), indicating a full mediation effect. These results confirm that purchase intention is shaped through emotional and social perceptions rather than voice type alone.

These findings confirm Hypotheses H9 to H12, highlighting the key role of emotions and social connection in AI-driven consumer decision-making. While previous models demonstrated the importance of trust and safety, Model 4 reveals that affective and relational factors also play a crucial role in shaping purchase intentions. Users do not simply evaluate voice assistants based on their functional attributes but also on their ability to evoke emotions and foster a sense of connection.

Figure 4.12. Model 4 results



Indirect Effects
Effect 1: Type of Voice → Emotional Response → Purchase Intention [0.03, 0.24]
Effect 2: Type of Voice → Social Attraction → Purchase Intention [0.05, 0.27]
Effect 3: Type of Voice → Emotional Response → Social Attraction → Purchase Intention [0.10, 0.34]

4.4.5. Model 5

This model examined how initial behavioral intention influences final purchase intention and whether voice type moderates this relationship. Following the methodological approach detailed in Section 4.3.3, a regression model with interaction terms was conducted to assess these effects.

The results confirmed that Behavior Intention significantly predicts PI ($\beta = 0.745$, $p < 0.001$), indicating that users with a strong initial intention to purchase via voice assistants are more likely to follow through with their decision. Moreover, the interaction term (Behavior Intention × Voice Type) was also significant ($F(2,896) = 9.615$, $p < 0.001$), demonstrating that the strength of this relationship varies across Siri, Google Assistant, and Alexa. These findings confirm H13, showing that voice type moderates the relationship between Behavioral Intention and final Purchase Intention.

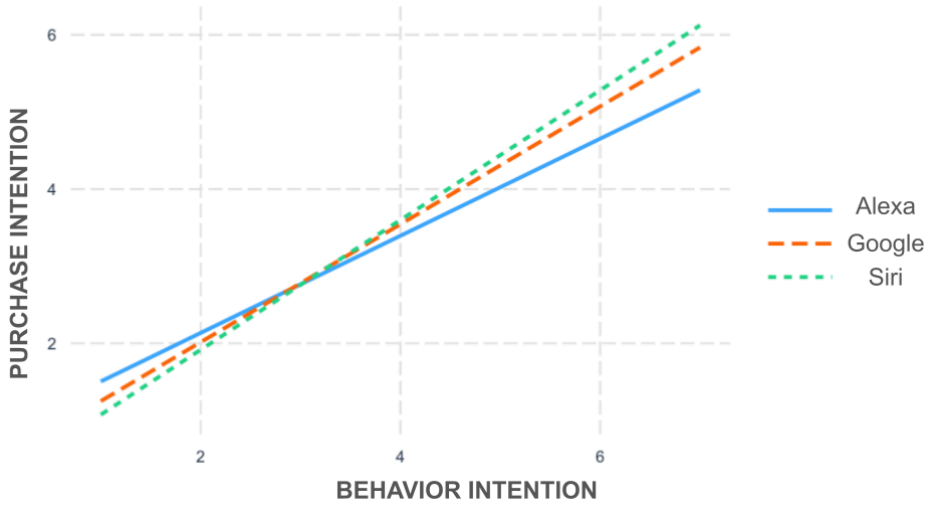
Post hoc comparisons revealed that the effect of Behavior Intention on PI is significantly weaker for Alexa users, while Siri and Google Assistant users exhibit stronger purchase consistency. Marginal slope estimations confirmed that Siri users (0.842, 95% CI: 0.774 - 0.909) and Google users (0.764, 95% CI: 0.695 - 0.834) maintain their initial purchase intentions more reliably than Alexa users (0.630, 95% CI: 0.561 - 0.698). The differences between Alexa and the other assistants were statistically significant, with Google

($p = 0.0189$) and Siri ($p < 0.0001$) showing stronger associations between initial and final purchase intention, while no significant difference was found between Google and Siri ($p = 0.2634$).

Notably, Siri exhibited the steepest slope, suggesting that users of this assistant display the highest consistency between their initial behavioral intention and their final purchase decision. This trend suggests that Siri's voice characteristics may enhance user confidence and reinforce purchase commitment, positioning it as the most effective assistant in translating behavioral intention into actual purchase behavior.

A visualization of this interaction effect (Figure 4.13) illustrates that Google and Siri display steeper slopes compared to Alexa, indicating a stronger influence of Behavioral Intention on Purchase Intention. The effect is particularly pronounced for Siri and Google, confirming that users of these assistants are more consistent in translating their initial purchase intention into final purchase decisions.

Figure 4.13. Visualization of interaction effect



4.4.6. Model 6

This model examined how Passive WOM influences Active WOM and whether voice type moderates this relationship. Following the methodological approach detailed in Section 4.3.3, a regression model with interaction terms was conducted to assess these effects.

The results showed that Passive WOM, measured using the indicator from Hedeman

(2021), had a mean of 2.59 and a standard deviation of 1.78. It significantly predicts Active WOM ($\beta = 0.384$, $p < 0.001$), indicating that users who received positive recommendations about a voice assistant before interacting with it were more likely to express an intention to recommend it themselves after the interaction. However, the interaction term (Passive WOM × Voice Type) was not statistically significant ($F(2,896) = 0.9937$, $p = 0.3706$), demonstrating that the strength of this relationship does not vary across Alexa, Google Assistant, and Siri.

Marginal slope estimates further support this conclusion, as all three assistants displayed a similar association between Passive and Active WOM. Siri users exhibited a slope of 0.480 (95% CI: 0.387 - 0.573), followed by Google users (0.431, 95% CI: 0.329 - 0.534), and Alexa users (0.384, 95% CI: 0.289 - 0.480). Post hoc comparisons did not reveal any significant differences between assistants, with all p-values far from conventional significance thresholds (Alexa vs. Google: $p = 0.7846$; Alexa vs. Siri: $p = 0.3364$; Google vs. Siri: $p = 0.7714$).

These findings suggest that the transition from receiving recommendations to intending to recommend the assistant is consistent across voice types, with no meaningful variation among assistants. As a result, Passive WOM influences Active WOM similarly, regardless of whether the voice assistant is Alexa, Google, or Siri. Given the absence of significant moderation effects, Hypothesis 14 is not supported. This indicates that voice type does not alter the strength of the Passive WOM–Active WOM relationship, and further analysis should explore other potential factors that may shape WOM conversion.

4.4.7. Model 7

This model examined whether voice type moderates the relationship between users' intention to recommend the voice assistant and final purchase intention, assessing how the willingness to advocate for a specific assistant influences purchase behavior differently across Siri, Google Assistant, and Alexa. As outlined in Section 4.3.3, a regression model with interaction terms was applied to test these effects.

The results confirmed that Active WOM significantly predicts Purchase Intention ($\beta = 0.704$, $p < 0.001$), indicating that higher WOM engagement leads to a stronger likelihood of purchasing. Moreover, the interaction term (Active WOM × Voice Type) was also significant ($F(2,896) = 7.296$, $p = 0.0007$), suggesting that the strength of Active WOM's influence on

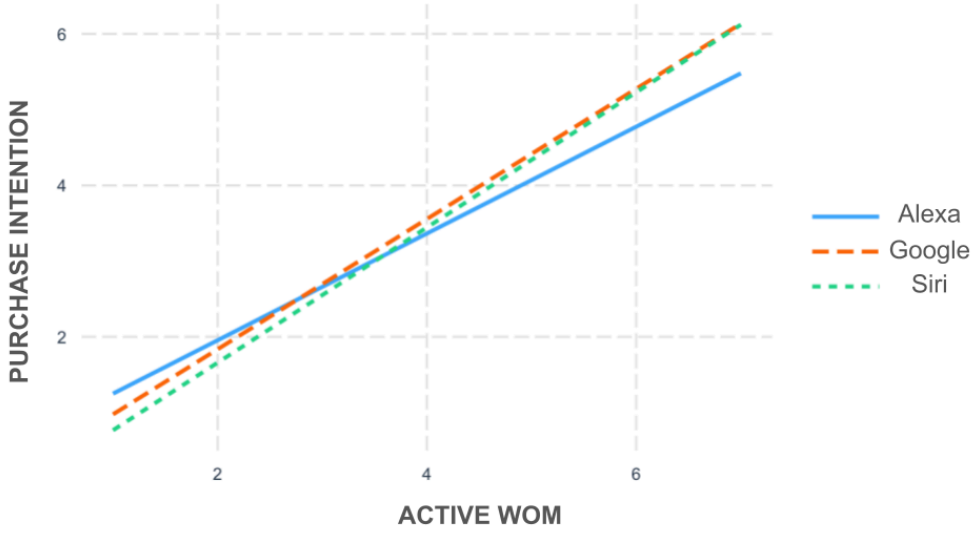
purchase intent varies depending on the assistant. These findings confirm Hypothesis H15, showing that voice type moderates the relationship between Active WOM and purchase intention.

The regression coefficients revealed that Active WOM has a weaker impact on purchase intention for Alexa users, whereas Google and Siri users exhibit a stronger relationship between Active WOM and Purchase Intention. Marginal slope estimations confirmed that Active WOM's effect is strongest for Siri users (0.892, 95% CI: 0.820 - 0.964), followed by Google users (0.860, 95% CI: 0.782 - 0.937), and weakest for Alexa users (0.705, 95% CI: 0.632 - 0.777). Figure 4.14 illustrates this interaction, showing that Google and Siri users display steeper slopes, reinforcing the idea that Active WOM plays a stronger role in their purchase decisions, while Alexa users show a weaker link between WOM and final purchase intention.

Post hoc comparisons further supported these findings, showing that after listening to the assistants, Google's Active WOM effect is significantly stronger than Alexa's ($p = 0.0117$), and Siri's WOM effect is also significantly stronger than Alexa's ($p = 0.0010$). While no significant difference was found between Google and Siri ($p = 0.8167$), the numerical advantage of Siri's slope suggests that, although weakly, Siri users might translate WOM into purchase intention more effectively than Google users. This aligns with prior findings suggesting that Siri's higher human-likeness and engagement ratings may provide it with a slight persuasive edge.

These results emphasize that voice assistants shape the effectiveness of WOM in consumer decision-making, with Siri and Google displaying the strongest WOM-to-purchase conversion, while Alexa users exhibit a weaker effect. Although statistical significance between Siri and Google was not reached, the observed trend suggests that Siri users may be slightly more receptive to WOM in driving final purchase decisions.

Figure 4.14. Visualization of interaction effect



To synthesize the key findings of our analyses, Table 4.2 presents a structured overview of the direct, mediated, and moderated effects observed across the different models. The table categorizes each effect based on whether the relationship was directly significant, mediated through one or more variables, or moderated by voice type or other factors.

Tabla 4.2. Voice effects on purchase intention

Effect Type	Direct Effect	Mediator	Moderator
Type of Voice → Purchase Intention	No	-	-
Type of Voice → Human-like voice → Purchase Intention	No	Yes	-
Type of Voice → Safety → Trust → Purchase Intention	No	Yes	-
Type of Voice → Emotional Response → Social Attraction → Purchase Intention	No	Yes	-
Behavior Intention → Purchase Intention	Yes	-	Yes
Passive WOM → Active WOM	Yes	-	No
Active WOM → Purchase Intention	Yes	-	Yes

4.5. Conclusions and discussion

Understanding how voice assistants influence purchase intention is essential for advancing AI-driven consumer behavior research.

This study aimed to analyze the key variables that explain purchase intention through voice assistants, with a particular focus on the impact of voice on users. To achieve this, we examined three types of voice assistants: Alexa, Google Assistant, and Siri. The analysis considered both the direct effect of voice on purchase intention and potential indirect effects through mediating variables. Additionally, the study explored the moderating role of voice type in shaping the relationship between previously identified factors in the literature that contribute to explaining purchase intention.

These findings build upon prior research demonstrating that human-like AI enhances social presence, engagement, and decision-making in digital commerce (Huh et al., 2023; Patrizi et al., 2021).

Model 1 tested the direct effect of voice type on purchase intention but found no significant relationship, suggesting that voice characteristics alone do not directly influence purchase intention. This aligns with previous research emphasizing that AI effectiveness is shaped more by user perceptions and psychological processing than by technical attributes alone (Tschopp & Sassenberg, 2024). Despite the absence of a direct effect, this study confirms that voice type significantly influences other key variables that help explain purchase intention, indirectly impacting the decision-making process. Model 2 demonstrated that perceived human-likeness fully mediates the relationship between voice type and purchase intention, with Siri being perceived as the most human-like voice. More expressive and natural-sounding voices enhanced perceptions of human-likeness, which, in turn, increased purchase likelihood, supporting the Computers Are Social Actors (CASA) framework (Nass & Moon, 2000). These findings expand upon prior research showing that human-like AI enhances social presence, engagement, and decision-making in digital commerce (Huh et al., 2023; Patrizi et al., 2021).

While human-likeness fosters trust (Schroeder et al., 2017), it is not sufficient on its own—users must also perceive the assistant as reliable, safe, and socially engaging for trust to translate into actual purchases. Model 3 introduced safety and trust as mediators, revealing

that expressive and natural-sounding assistants enhance perceptions of security, which in turn increase trust and purchase intention. Once again, Siri emerges as the voice that inspires the greatest sense of security. These findings align with (McLean & Osei-Frimpong, 2019), who emphasized that AI reliability is a key determinant of successful digital transactions. The full mediation effect (Safety → Trust → Purchase Intention) underscores that for AI-driven commerce to be effective, users must first feel secure before they can trust the system. This highlights the need for trust-enhancing strategies, such as transparent privacy policies, reliability cues, and user feedback mechanisms, to strengthen consumer confidence in AI-powered transactions.

Model 4 expanded on previous findings by incorporating emotional response and social attraction as mediators. Once again, Siri stands out for its ability to convey emotions, which has a stronger impact on purchase intention compared to Alexa or Google. Voice expressiveness enhances emotional responses, which, in turn, increase social attraction, ultimately leading to higher purchase intention. This reinforces the idea that AI assistants are not merely transactional tools but also affective and social agents that shape user experiences and purchasing decisions. The mediation pathway (Emotional Response → Social Attraction → Purchase Intention) highlights the importance of affective and relational factors in AI-driven commerce, aligning with Schroeder et al. (2017). AI that elicits emotions and fosters social presence enhances user engagement and trust, making interactions more meaningful. These findings suggest that AI commerce strategies should go beyond efficiency and accuracy to prioritize emotional and social engagement. Customizing AI voices to match user preferences and integrating voice-based storytelling could further strengthen emotional connections with consumers, enhancing their overall experience.

While Models 1–4 demonstrated that human-likeness, trust, emotional engagement, and social attraction are key drivers of purchase intention, Models 5, 6, and 7 introduced moderation effects of voice type, providing deeper insight into how these processes vary across voice type. These findings underscore the dynamic nature of AI-driven consumer behavior and emphasize the need for personalized, context-aware strategies in voice commerce to enhance user experience and engagement.

Model 5 explored how the relationship between initial and final purchase intention varies across voice assistants. Once again, Siri stands out for their stronger purchase consistency compared to Alexa, indicating that users of these assistants are more likely to

follow through with their initial purchase intentions. While behavioral intention remains a key predictor of purchase decisions, Alexa users show a weaker connection between their initial intent and final choice, suggesting that additional factors may influence their decision-making process. This reinforces the idea that perceived reliability and persuasiveness of AI interfaces play a crucial role in consumer decision stability, as suggested by Huh et al. (2023). These findings highlight the importance of tailoring AI commerce strategies to different assistant ecosystems. For Google and Siri, reinforcing initial purchase intent through trust-building mechanisms can further strengthen decision consistency, while Alexa users may require contextual incentives or adaptive engagement strategies to enhance conversion rates.

Model 6 examined how Passive WOM influences Active WOM and whether this relationship varies across voice assistants. The results confirm that users who receive positive recommendations about a voice assistant before interacting with it are more likely to recommend the assistant themselves after the interaction, reinforcing the idea that perceptions formed through WOM play a key role in user advocacy. However, this effect does not differ across Alexa, Google, and Siri, suggesting that the process of transitioning from receiving recommendations to actively recommending the assistant remains consistent across voice types. This aligns with research emphasizing that user advocacy is primarily shaped by personal experience rather than the specific AI interface. These findings indicate that Passive WOM drives Active WOM similarly across all assistants, highlighting the robustness of WOM influence in voice commerce.

The findings from Models 5 and 6 align with Expectation Confirmation Theory (ECT), which posits that users reinforce experiences that match their pre-existing beliefs (Kim et al., 2021). When their interactions with a voice assistant contradict initial expectations, users reassess their perceptions before expressing Active WOM (Reber et al., 2004).

Building on Model 6, Model 7 examined whether voice type moderates the link between Active WOM and final purchase intention. While Active WOM significantly influenced purchase intent across all assistants, the effect was stronger for Siri and Google Assistant users than for Alexa users. These findings align with research suggesting that certain AI interfaces amplify social influence effects, making users more receptive to peer recommendations (Huh et al., 2023).

Model 7 explored whether voice assistant type moderates the relationship between Active WOM and purchase intention, revealing that while Active WOM significantly predicts purchase decisions across all assistants, its effect is strongest for Siri users, followed by Google users, and weakest for Alexa users. This reinforces the idea that certain AI interfaces enhance social influence effects, making users more receptive to peer recommendations, as suggested by Huh et al. (2023). Siri emerges as the voice with the greatest influence on purchase intention, both directly and indirectly, confirming its impact across multiple psychological and social factors. The interaction analysis further supports this, aligning with prior findings on Siri's superior human-likeness and friendliness (Figure 4.8), showing that Siri users display the steepest slopes, meaning that their purchase decisions are the most influenced by WOM engagement, followed by Google users, while Alexa users show the weakest link between recommendations and final purchase intention.

The findings of this study provide compelling evidence that a voice assistant's vocal characteristics are not merely a functional aspect but a defining element in enhancing the user experience, shaping consumer trust, sense of security, and purchasing behavior. Among the three assistants examined, Siri consistently stands out as the voice that establishes the strongest psychological and emotional connection with users, leading to greater trust, security, and persuasiveness. More than just an interface, a well-designed voice becomes a social actor, aligning with the Computers Are Social Actors (CASA) framework (Nass & Moon, 2000), capable of influencing consumer decisions.

In this light, voice assistants should not only focus on usability but also prioritize human-like, warm, and expressive vocal traits to optimize the user experience, engagement, and commercial success. Brands and developers that fail to invest in refining these characteristics may struggle to build long-term trust and influence over their users. As AI-driven commerce continues to expand, enhancing and fine-tuning voice assistant design will be crucial for maintaining a competitive edge in an increasingly conversational and interactive market.

Future research should further explore the nuances of Siri-like voice attributes that most effectively drive human-likeness, emotional expression, and trust, ensuring that digital voices evolve to enhance interaction experiences not only with the virtual assistants we know today but also with any voice-enabled device used in customer interactions.

4.5.1. Implications for theory

This study advances Social Response Theory (SRT), AI trust formation, and word-of-mouth (WOM) research, demonstrating that voice assistants actively shape user engagement rather than serving as neutral intermediaries.

Models 2-4 confirm that human-likeness, trust, and emotional connection fully mediate the relationship between voice type and purchase behavior, reinforcing SRT by showing that users apply human-like social norms to AI interactions (Nass & Moon, 2000). However, this study refines SRT by demonstrating that AI engagement varies based on perceived safety, trust, and emotional connection, rather than voice expressiveness alone. Siri consistently elicits stronger trust and emotional engagement, followed by Google, while Alexa remains the least effective in fostering user connection.

Model 3 highlights that AI trust is not an immediate outcome but a sequential process, where users must first perceive the assistant as safe before trust can develop and lead to purchases. Model 4 further supports the idea that trust is multidimensional, combining rational assessments (competence, reliability) with emotional perceptions (warmth, engagement) (Huh et al., 2023). These findings refine trust formation theories by emphasizing that AI credibility is as much about emotional resonance as it is about perceived accuracy.

Models 5-7 extend these insights by exploring how voice type moderates key behavioral relationships. Model 5 challenges traditional consumer behavior models by showing that behavioral intention does not automatically translate into purchase intention—this relationship depends on the AI interface itself. Siri and Google reinforce purchase stability, while Alexa users are more likely to reconsider their decisions, suggesting that AI reliability plays a moderating role in consumer consistency.

Model 6 contributes to WOM research by demonstrating that Passive WOM influences Active WOM similarly across all assistants, reinforcing that WOM advocacy is a stable process unaffected by voice type. However, Model 7 reveals that Active WOM's impact on purchase decisions is significantly stronger for Siri, followed by Google, while Alexa users show the weakest WOM-driven purchasing behavior. This aligns with theories on AI as a social persuasion agent, supporting the idea that AI can amplify or weaken social influence depending on its perceived credibility and persuasiveness (Keller & Fay, 2012).

These findings suggest that AI trust, social influence, and persuasion are dynamic processes shaped by psychological and emotional mechanisms rather than purely functional attributes. By demonstrating how AI characteristics mediate and moderate consumer decision-making, this study refines existing theories on human-AI interaction, trust formation, and digital persuasion.

4.5.2. Implications for practice

This study provides actionable insights for businesses, AI developers, and marketers leveraging voice assistants in digital commerce. Rather than serving as passive tools, voice assistants actively shape consumer trust, engagement, and purchase behavior, making it crucial for companies to optimize AI-driven interactions beyond basic functionality.

Trust formation is a key driver of AI adoption, but as Model 3 confirms, it does not emerge instantly—it is built through perceptions of safety, human-likeness, and emotional engagement. To maximize consumer trust, AI-powered voice assistants should prioritize natural intonation, expressiveness, and conversational fluidity, reinforcing human-likeness without triggering discomfort (Tschopp & Sassenberg, 2024). Siri, which consistently fosters the strongest trust and engagement, sets the benchmark for effective AI design, while Google follows closely and Alexa lags behind. Since safety perceptions are a prerequisite for trust formation, businesses should enhance transparency in AI interactions, clearly communicating privacy protections, encryption measures, and explicit consent options to build user confidence.

AI-driven marketing strategies should be tailored to different consumer segments and product categories, while also prioritizing the optimization of voice assistant characteristics to strengthen user relationships and purchase intentions (Models 5 and 6). Model 5 confirms that prior purchase intentions are reinforced when the assistant's voice is perceived as more human-like and engaging. Therefore, companies should focus on refining voice characteristics to resemble those that elicit stronger emotional connections, such as the tones used by Siri, which enhance trust and engagement.

Models 6 and 7 confirm that voice assistants moderate the impact of WOM on purchase behavior, with effectiveness varying across platforms. Siri, followed by Google, amplifies WOM's impact on purchase intent, making these assistants ideal for AI-driven

social proof strategies such as voice-generated product recommendations, interactive referral programs, and AI-driven influencer endorsements. In contrast, Alexa users exhibit a weaker WOM-to-purchase link, suggesting that businesses should reinforce purchase confidence through personalized incentives, AI-driven discounts, time-sensitive promotions, and reminder-based engagement strategies.

To remain competitive in AI-driven commerce, businesses must invest in adaptive AI voice technologies that dynamically tailor interactions based on user sentiment, preferences, and shopping behavior. This includes real-time tone adjustments, speech customization for different purchase types, and user-personalized voice settings. Siri's ability to evoke emotional engagement and trust positions it as the optimal model for AI-driven consumer influence, while Google provides strong but slightly less effective engagement, and Alexa requires additional enhancements to improve its persuasive impact.

This study confirms that voice assistants are not merely functional tools but persuasive agents shaping consumer trust, engagement, and decision-making. Businesses must move beyond treating AI as a transactional system and instead integrate psychological and social influence mechanisms into AI-driven marketing. By leveraging trust-building strategies, optimizing WOM influence, and implementing adaptive personalization, companies can create more engaging, persuasive, and consumer-centric AI-driven shopping experiences.

4.5.3. Limitations and future research

While this study provides valuable insights into the role of voice assistants in shaping consumer trust and purchase intention, several limitations should be acknowledged.

First, the study focused on three widely used voice assistants (Siri, Google Assistant, and Alexa), which may limit the generalizability of the findings. Future research should explore a broader range of AI voices, including those with different accents, speech cadences, or gender variations, to assess their impact on consumer behavior.

Second, trust was measured at a single point in time, yet AI trust evolves through repeated interactions. Future studies could adopt longitudinal designs to examine how continued AI usage influences trust stability, habit formation, and sustained purchase behavior.

Third, the study relied on self-reported data, which may be influenced by social desirability bias or user expectations. Incorporating behavioral experiments, biometric tracking with neuromarketing studies (e.g., eye-tracking, heart rate monitoring), or real-world AI purchase data could provide a more objective understanding of user responses. Additionally, cultural and demographic factors may shape AI perceptions differently. Future research should explore cross-cultural differences in AI trust and voice preferences, particularly in regions with varying levels of technology adoption and privacy concerns.

Additionally, this study used the Spanish (Spain) versions of the voice assistants, which may influence the findings. Voice perception can vary across languages due to phonetic, prosodic, and cultural differences, potentially shaping trust, emotional engagement, and purchase intention in distinct ways. Future research should examine whether these effects hold across other languages and linguistic contexts, exploring how variations in speech synthesis impact consumer responses in different markets.

Moreover, purchase intention is likely influenced not only by the characteristics of the voice assistant but also by external factors, such as trust in the underlying shopping platform. For example, Alexa's association with Amazon or Google Assistant's integration with Google Shopping may shape user confidence in transactions beyond the assistant's voice itself. Additionally, the degree of integration between the assistant and the shopping service may vary across platforms, potentially influencing ease of use, perceived reliability, and overall purchase likelihood. Future research should examine how the interaction between voice assistant characteristics and platform trust impacts consumer behavior in voice commerce.

Finally, as AI continues to evolve, future studies should explore adaptive AI voices that dynamically adjust tone and expressiveness based on user sentiment and interaction context. Investigating whether real-time AI personalization enhances trust and purchase conversion could provide critical insights for the next generation of voice commerce strategies.

4.6. Conclusions

Voice assistants are more than functional tools; they actively shape consumer trust, engagement, and decision-making in AI-driven commerce. This study confirms that VAs do not directly influence purchase behavior but rather operate through psychological mediators

such as trust, safety, emotional engagement, and social attraction. These findings refine Social Response Theory by demonstrating that users apply human-like social norms to AI, but their engagement depends on cognitive and emotional evaluations rather than voice naturalness alone.

Among these mediators, trust emerged as the strongest predictor of purchase intention, reinforcing previous research on AI credibility and social presence. However, trust is not formed in isolation—it is preceded by safety perceptions, which determine whether users feel secure enough to engage in AI-assisted purchases. Siri consistently fosters the highest trust and emotional engagement, followed by Google, while Alexa users display greater variability in decision-making. These results suggest that Siri and Google users form stronger initial purchase intentions, whereas Alexa users are more susceptible to external influences like promotions or real-time interaction quality.

The study also underscores the role of word-of-mouth in AI commerce, showing that Passive WOM consistently translates into Active WOM across all assistants. However, the impact of the intention to recommend a voice assistant on purchase intention is stronger for Siri and Google users, whereas Alexa users show a weaker relationship. This suggests that users who are more willing to recommend an assistant also tend to be more satisfied with it, reinforcing the idea that WOM serves as a proxy for loyalty.

Finally, these findings highlight that VAs do not directly drive consumer decisions but rather influence key psychological and social processes that shape purchase intention. The effectiveness of trust, safety, emotional engagement, and WOM differs across assistants, underscoring the need for tailored AI voice commerce strategies. Siri stands out as the most human-like voice, effectively conveying emotions and providing a greater sense of security to users, which ultimately leads to a higher intention to use it as a shopping assistant.

4.7. References

Aguirre-Urreta, M. I., Marakas, G. M., & Ellis, M. E. (2013). Measurement of composite reliability in research using partial least squares: Some issues and an alternative approach. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, *44*(4), 11-43. https://doi.org/10.1145/2544415.2544417

Baranger, D. A., Finsaas, M. C., Goldstein, B. L., Vize, C. E., Lynam, D. R., & Olino, T. M. (2023). Tutorial: Power analyses for interaction effects in cross-sectional regressions. *Advances in Methods and Practices in Psychological Science*, *6*(3), 25152459231187531.

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, *1*, 71-81.

Bawack, R. E., Wamba, S. F., & Carillo, K. D. A. (2021). Exploring the role of personality, trust, and privacy in customer experience performance during voice shopping: Evidence from SEM and fuzzy set qualitative comparative analysis. *International Journal of Information Management*, *58*, 102309. https://doi.org/10.1016/j.ijinfomgt.2021.102309

Berger, J. (2013). *Contagious: How to build word of mouth in the digital age*. Simon and Schuster.

Berger, J. (2014). Word of mouth and interpersonal communication: A review and directions for future research. *Journal of consumer psychology*, *24*(4), 586-607.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, *25*(1), 49-59.

Casado-Aranda, L.-A., Dimoka, A., & Sánchez-Fernández, J. (2019). Consumer processing

of online trust signals: A neuroimaging study. *Journal of Interactive Marketing*, *47*(1), 159-180.

Casado-Aranda, L.-A., Liébana-Cabanillas, F., & Sánchez-Fernández, J. (2018). A neuropsychological study on how consumers process risky and secure E-payments. *Journal of Interactive Marketing*, *43*(1), 151-164.

Charalambous, G., Fletcher, S., & Webb, P. (2016). The Development of a Scale to Evaluate Trust in Industrial Human-robot Collaboration. *International Journal of Social Robotics*, *8*(2), 193-209. https://doi.org/10.1007/s12369-015-0333-8

East, R., Hammond, K., & Lomax, W. (2008). Measuring the impact of positive and negative word of mouth on brand purchase probability. *International journal of research in marketing*, *25*(3), 215-224.

Fornell, C., & Larcker, D. F. (1981). Structural Equation Models with Unobservable Variables and Measurement Error: Algebra and Statistics. *Journal of Marketing Research*, *18*(3), 382-388. https://doi.org/10.2307/3150980

Grewal, D., Guha, A., Schweiger, E., Ludwig, S., & Wetzels, M. (2022). How communications by AI-enabled voice assistants impact the customer journey. *Journal of Service Management*, *33*(4/5), 705-720.

Hair, J. (2009). Multivariate Data Analysis. *Faculty and Research Publications*. https://digitalcommons.kennesaw.edu/facpubs/2925

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.

Hedeman, J. (2021, enero). *Hey Siri, let's go shopping! : A study into the factors influencing Dutch consumers' intention to use a voice assistant for online shopping* [Info:eu-repo/semantics/masterThesis]. University of Twente. http://essay.utwente.nl/85619/

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the internet? *Journal of interactive marketing*, *18*(1), 38-52.

Henseler, J., Hubona, G., & Ray, P. A. (2016). Using PLS path modeling in new technology research: Updated guidelines. *Industrial Management & Data Systems*, *116*(1), 2-20. https://doi.org/10.1108/IMDS-09-2015-0382

Huh, J., & Kim, H. Y. (2022). The Uses and Gratifications Model of Voice Shopping. *International Textile and Apparel Association Annual Conference Proceedings*, *78*(1).

Huh, J., Whang, C., & Kim, H.-Y. (2023). Building trust with voice assistants for apparel shopping: The effects of social role and user autonomy. *Journal of Global Fashion Marketing*, *14*(1), 5-19.

Jang, Y.-T. J., Liu, A. Y., & Ke, W.-Y. (2022). Exploring smart retailing: Anthropomorphism in voice shopping of smart speaker. *Information Technology & People*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/ITP-07-2021-0536

Kahneman, D. (2011). *Thinking, fast and slow*. macmillan.

Keerthi, S. (2024). Voice-activated e-commerce: Revolutionizing shopping experiences with personalization and seamless integration. *International Journal for Research in Applied Science and Engineering Technology*.

Keller, E. B., & Fay, B. (2012). *The face-to-face book: Why real relationships rule in a digital marketplace*. Simon and Schuster.

Kim, J., Lee, K., & Kim, S. (2021). Emotional influence of virtual assistant voices on user engagement and trust. *International Journal of Human-Computer Interaction*, *37*(8), 702-717.

Maduku, D. K., Rana, N. P., Mpinganjira, M., Thusi, P., Mkhize, N. H.-B., & Ledikwe, A. (2024). Do AI-powered digital assistants influence customer emotions, engagement

and loyalty? An empirical investigation. *Asia Pacific Journal of Marketing and Logistics*, *36*(11), 2849-2868.

McLean, G., & Osei-Frimpong, K. (2019). Hey Alexa … examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in Human Behavior*, *99*, 28-37. https://doi.org/10.1016/j.chb.2019.05.009

Moriuchi, E. (2019). Okay, Google!: An empirical study on voice assistants on consumer engagement and loyalty. *Psychology & Marketing*, *36*(5), 489-501. https://doi.org/10.1002/mar.21192

Nass, C., & Brave, S. (2005). *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. MIT Press.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of social issues*, *56*(1), 81-103.

Patrizi, M., Vernuccio, M., & Pastore, A. (2021). «Hey, voice assistant!» How do users perceive you? An exploratory study. *Sinergie Italian Journal of Management*, *39*(1), 173-192.

Pias, S. B. H., Huang, R., Williamson, D. S., Kim, M., & Kapadia, A. (2024). The Impact of Perceived Tone, Age, and Gender on Voice Assistant Persuasiveness in the Context of Product Recommendations. *Proceedings of the 6th ACM Conference on Conversational User Interfaces*, 1-15.

Poushneh, A. (2021). Humanizing voice assistant: The impact of voice assistant personality on consumers' attitudes and behaviors. *Journal of Retailing and Consumer Services*, *58*, 102283. https://doi.org/10.1016/j.jretconser.2020.102283

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior research methods*, *40*(3), 879-891.

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and social psychology review*, *8*(4), 364-382.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161-1178. https://doi.org/10.1037/h0077714

Rzepka, C., Berger, B., Koslow, A., & Hess, T. (2023). Voice Assistant, Buy Coffee Capsules!: Understanding the Determinants of Consumers' Intention to Use Voice Commerce. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, *54*(3), 137-159.

Schroeder, J., Kardas, M., & Epley, N. (2017). The humanizing voice: Speech reveals, and text conceals, a more thoughtful mind in the midst of disagreement. *Psychological science*, *28*(12), 1745-1762.

Schultz, C. D., & Paetz, F. (2023). Trust in Digital Voice Assistants: A Fundamental Determinant for Companies' and Customers' Engagement in Voice Commerce. *Marketing ZFP-Journal of Research & Management*, *45*(2).

Statista. (2023). *Number of digital voice assistants in use worldwide from 2019 to 2024 (in billions)*. https://www.statista.com/statistics/973815/worldwide-digital-voice-assistant-in-use/#:~:text=Number%20of%20digital%20voice%20assistants%20in%20use%20worldwide%202019%2D2024&text=In%202020%2C%20there%20will%20be,higher%20than%20the%20world's%20population.

Strathmann, C., Szczuka, J., & Krämer, N. (2020). She talks to me as if she were alive: Assessing the social reactions and perceptions of children toward voice assistants and their appraisal of the appropriateness of these reactions. *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, 1-8.

Sun, C., Shi, Z. (June), Liu, X., Ghose, A., Li, X., & Xiong, F. (2019). *The Effect of Voice AI on Consumer Purchase and Search Behavior* (SSRN Scholarly Paper 3480877). Social Science Research Network. https://doi.org/10.2139/ssrn.3480877

Tobon, S., & García-Madariaga, J. (2021). The influence of opinion leaders' ewom on online consumer decisions: A study on social influence. *Journal of Theoretical and Applied Electronic Commerce Research*, *16*(4), 748-767.

Tobon, S., Ruiz-Alba, J. L., & García-Madariaga, J. (2020). Gamification and online consumer decisions: Is the game over? *Decision Support Systems*, *128*, 113167.

Tschopp, M., & Sassenberg, K. (2024). The impact of human-AI relationship perception on voice shopping intentions. *Human-Machine Communication*, *8*, 101-117.

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly: Management Information Systems*, *36*(1), 157-178.

Wang, Y., & Yu, C. (2017). Social interaction-based consumer decision-making model in social commerce: The role of word of mouth and observational learning. *International Journal of Information Management*, *37*(3), 179-189.

Whang, C. (2018). *Voice shopping: The effect of the consumer-voice assistant parasocial relationship on the consumer's perception and decision making*. 173.

Zhu, Q., Chau, A., Cohn, M., Liang, K.-H., Wang, H.-C., Zellou, G., & Yu, Z. (2022). Effects of emotional expressiveness on voice chatbot interactions. *Proceedings of the 4th conference on conversational user interfaces*, 1-11.

4.8. Appendix

Table 4.3. Scale

Item	Scales
Purchase Intention - Adapted from Bawack et al. (2021)	
PUR1	It is likely that I will use this VA to shop online
PUR2	I would use this VA to shop online in the near future
Voice characteristics - Adapted from Whang (2018)	
The assistant's voice seemed to me…	
VC1	Robotic (1) - Human (7)
VC2	Irritating (1) - Friendly (7)
Emotional Response - Adapted from Russell (1980)	
The assistant's voice seemed to me…	
ER1	Very negative (1) - Very positive (7)
This assistant's voice has made me feel…	
ER2	Very relaxed (1) - Very nervous (7)
Human Like Voice - Adapted from Patrizi et al. (2021)	
HLV1	The assistant's voice sounds human
HLV2	The quality of his voice seems human
HLV3	The tone of this VA's voice seems human
Trust - Adapted from Charalambous et al. (2016)	
TRU1	The assistant's voice makes me feel comfortable
TRU2	The assistant's voice would make me trust them
TRU3	I would feel comfortable interacting with this VA
Active Word of mouth - Adapted from Moriuchi (2019)	
AWOM1	I would recommend friends and family to buy with this VA
AWOM2	I would encourage my close circle to voice shopping by using a VA with this voice

Passive Word of mouth - Adapted from Hedeman (2021)	
PWOM1	I have heard positive opinions about online shopping with virtual assistants
Social attraction - Adapted from McLean & Osei-Frimpong (2019)	
SATR1	I think this VA and I could be friends
SATR2	I could have a good time with this VA
Perceived Safety - Adapted from McLean & Osei-Frimpong (2019)	
PSAF1	I feel safe using VA
PSAF2	I feel that VAs are a safe technology
Behavior Intention - Adapted from Bawack et al. (2021)	
BE1	In the future I will continue to use VA in my daily life (eg.: music, weather, shopping…)
BE2	In the future I will continue to use VA to search for purchasing information
BE3	In the future I will use VA to shop online

CONCLUSIONES

5.1. Conclusiones

Esta tesis doctoral ha tenido como objetivo comprender cómo optimizar la experiencia del consumidor en el comercio electrónico, centrándose en la interacción con asistentes virtuales de voz. Más allá de analizar la funcionalidad de estos sistemas, se ha estudiado su impacto en la percepción de confianza, seguridad y transmisión de emociones, considerando cómo los consumidores procesan e interpretan las señales vocales en entornos de compra asistida por inteligencia artificial (Fakhimi et al., 2023).

Con el fin de profundizar en estos aspectos, la presente investigación ha abordado el comercio por voz desde una perspectiva centrada en el usuario, explorando cómo la tipología de voz y la antropomorfización afectan la intención de compra. En particular, se han analizado los efectos de la voz de tres tipos de asistentes virtuales (Alexa, Google y Siri) en la predisposición del consumidor a interactuar y realizar compras a través de ellos (Huh et al., 2023). Los resultados han mostrado que la percepción de humanidad y presencia social son factores determinantes en la confianza del usuario, alineándose con estudios previos sobre la interacción con IA y la teoría *Computers Are Social Actors* (CASA) (Nass & Moon, 2000).

Además de los aspectos técnicos y del diseño vocal, la confianza del usuario en los asistentes de voz se ve influenciada por factores como la percepción de presencia social y la seguridad en la interacción. Se ha demostrado que una voz cálida y natural incrementa la confianza y reduce la percepción de riesgo, mientras que una voz excesivamente humanizada puede generar rechazo, en línea con la hipótesis del *Uncanny Valley* (Deshpande et al., 2023; Mori, 1970). Este equilibrio entre humanización y funcionalidad es clave para mejorar la adopción del comercio por voz y garantizar una experiencia satisfactoria (Chen & Park, 2021).

Desde el punto de vista teórico, esta tesis doctoral contribuye a la literatura sobre interacción humano-ordenador, al ofrecer una visión integral de cómo la voz impacta la percepción del usuario en entornos de compra online. Se han aplicado modelos como el *Technology Acceptance Model* (TAM) (Davis, 1989) y la *Uses and Gratifications Theory* (UGT) (Katz et al., 1973) para explicar la aceptación de esta tecnología, complementándolos con estudios sobre antropomorfización en asistentes virtuales y la *Social Response Theory* (SRT) (Nass & Moon, 2000).

La evolución del comercio electrónico y la inteligencia artificial plantea el reto de diseñar interacciones más naturales y humanas. La creciente automatización de la experiencia de compra ha impulsado nuevas estrategias para fortalecer la relación entre los usuarios y los asistentes de voz, con el objetivo de aumentar la confianza y facilitar la adopción del canal vocal. En este contexto, las empresas y los diseñadores de asistentes virtuales deben apostar por enfoques que optimicen la experiencia del consumidor, asegurando un equilibrio entre la eficiencia tecnológica y la cercanía comunicativa (Mari et al., 2024). La adaptación a distintos perfiles de usuario y la personalización del tono y la expresividad vocal serán elementos clave para fomentar la aceptación y el uso continuado de estas herramientas en el futuro del comercio digital (Sun & Liang, 2023).

En respuesta a las prioridades de investigación del Marketing Science Institute (MSI) 2022-2024 y a las brechas existentes en la literatura sobre comercio por voz, esta tesis ha desarrollado cuatro estudios empíricos que contribuyen a una mejor comprensión del comportamiento del consumidor en la interacción con asistentes virtuales.

El **primer estudio** analiza los factores que influyen en la adopción del comercio por voz a través del Technology Acceptance Model y la Uses and Gratifications Theory. Se examina cómo la percepción de utilidad, seguridad y disfrute impacta la intención de uso de asistentes de voz en compras online. Los resultados confirman que la utilidad percibida y el disfrute son determinantes clave en la adopción de estos sistemas, mientras que la seguridad percibida sigue siendo una barrera importante. Esto resalta la necesidad de estrategias que fortalezcan la confianza del usuario en la tecnología para facilitar su integración en los hábitos de compra.

El **segundo estudio** evalúa el impacto de la antropomorfización vocal en la percepción de seguridad y confianza, y cómo estos factores influyen en la disposición del usuario a comprar mediante asistentes de voz. Se estructura en dos partes: primero, analiza cómo la humanización de la voz, independientemente del tipo específico, mejora la percepción de seguridad; segundo, examina cómo esta percepción afecta la intención de compra. Los resultados indican que la presencia social percibida juega un papel clave como mediador entre la humanización de la voz y la confianza del usuario. Aunque no se identifican diferencias significativas entre tipos de voz, el simple hecho de introducir voz humanizada mejora la percepción del sistema. Estos hallazgos aportan tanto a la

investigación académica como al desarrollo de asistentes de voz más efectivos para el comercio digital.

El **tercer estudio** analizó el papel de los asistentes virtuales (Siri, Google Assistant y Alexa) en la transmisión de emociones básicas y su impacto en la confianza y percepción de humanización. Los resultados muestran que las voces pueden comunicar emociones que generan una respuesta emocional en el usuario, influyendo en la percepción de humanización y confianza. Siri destacó por su capacidad para transmitir emociones positivas, generando una mayor conexión emocional en comparación con Google Assistant y Alexa. Se comprobó que las emociones positivas aumentan la calidez y confianza percibida en el asistente, mientras que las negativas reducen la percepción de humanidad y la confianza del usuario. Además, se evaluó el impacto de la valencia del mensaje, concluyendo que los mensajes con carga emocional positiva refuerzan la percepción de confianza y humanización. Estos hallazgos subrayan la importancia del diseño vocal en asistentes de voz para mejorar la experiencia del usuario y fortalecer la identidad de marca.

El **cuarto estudio** analizó cómo el tipo de voz en los asistentes virtuales influye en la percepción de seguridad, la confianza y la intención de compra. Se evaluaron distintos estilos vocales para determinar su impacto en la disposición del usuario a utilizar el asistente para realizar compras mediante comandos de voz. Los resultados confirman que las voces más naturales y expresivas generan mayor seguridad y confianza, lo que incrementa la intención de compra. Se identificó que la seguridad percibida media la relación entre el tipo de voz y la confianza, y que la respuesta emocional y la presencia social también influyen en la aceptación del asistente como canal de compra. Entre los asistentes evaluados, Siri destaca como la voz más efectiva, al generar mayor confianza y conexión con el usuario, mientras que las voces más robóticas o sintéticas reducen la persuasión, limitando la intención de compra en el comercio digital.

Estos resultados sugieren que el diseño vocal de los asistentes virtuales debe ajustarse a las expectativas del consumidor para optimizar la experiencia de compra. A pesar de ciertas variaciones en la percepción de la voz según el perfil del usuario, se confirma que una voz bien diseñada puede aumentar la confianza y favorecer la conversión en entornos de comercio por voz. Además, se identificó que la percepción de humanidad en la voz del asistente desempeña un papel crucial en la interacción usuario-tecnología. Aquellos asistentes con voces más naturales y expresivas fomentaron un mayor grado de socialización, lo que

refuerza la intención de compra y el boca a boca electrónico.

5.2. Implicaciones teóricas

En línea con el objetivo de obtener un conocimiento más profundo sobre la experiencia del consumidor en el comercio por voz, los cuatro estudios que componen esta tesis presentan importantes implicaciones teóricas.

El primer capítulo contribuye a la literatura sobre adopción de tecnologías en el comercio electrónico, al examinar cómo la percepción de utilidad, gratificación y seguridad influyen en la intención de los consumidores de utilizar asistentes de voz para compras en línea. Basado en modelos como el TAM y la UGT, este estudio resalta la importancia de la seguridad percibida como una barrera clave para la adopción de esta tecnología.

Estos hallazgos amplían el conocimiento sobre factores que afectan la adopción del comercio por voz, alineándose con las investigaciones que analizan la confianza y la percepción de riesgo en la interacción humano-ordenador. La contribución teórica de este estudio radica en la integración de modelos de adopción de tecnología con el análisis de la seguridad percibida, proporcionando un marco más completo para comprender cómo los consumidores aceptan esta modalidad de compra.

El segundo capítulo contribuye a la literatura sobre antropomorfización en inteligencia artificial al examinar el impacto de los asistentes de voz con características humanas en la percepción de seguridad y confianza del usuario. Basándose en la teoría CASA, los hallazgos confirman que los usuarios tienden a percibir asistentes virtuales más humanizados como agentes sociales confiables, lo que impacta directamente su disposición a utilizar la tecnología para realizar compras.

Una de las principales aportaciones teóricas de este estudio es la identificación de la seguridad percibida como un mediador clave en la adopción del comercio por voz, integrando elementos del TAM con variables de antropomorfización. Se demuestra que la confianza en los asistentes virtuales no solo se basa en factores de funcionalidad, sino que está influenciada por cualidades humanas percibidas, como la calidez de la voz y la sensación de presencia social.

Además, este estudio amplía el debate sobre la teoría *Communication Privacy*

Management (CPM) al sugerir que la percepción de humanidad en los asistentes puede reducir la percepción de riesgo y facilitar la divulgación de información personal en interacciones comerciales. Estos resultados refuerzan la idea de que la relación usuario-asistente no es meramente transaccional, sino que se desarrolla dentro de un marco de confianza que puede ser modulado por la percepción de antropomorfización.

En términos más amplios, este capítulo aporta evidencia empírica sobre la importancia de los factores sociales y emocionales en la adopción de asistentes de voz, contribuyendo a la literatura sobre la interacción humano-ordenador y ofreciendo una base teórica para futuros estudios sobre el impacto de la humanización en la aceptación de tecnologías conversacionales.

El tercer capítulo contribuye al conocimiento sobre la confianza en los asistentes virtuales de voz, explorando cómo la emocionalidad en la voz y la valencia del mensaje afectan la percepción de humanidad y credibilidad del sistema. A nivel teórico, este estudio profundiza en la intersección entre la teoría de la respuesta emocional y el concepto de antropomorfismo en la inteligencia artificial, demostrando que la transmisión de emociones en la voz de un asistente virtual puede fortalecer su percepción como entidad confiable y socialmente presente.

Una de las principales contribuciones es la validación de la teoría de la presencia social en el contexto de asistentes de voz, sugiriendo que la expresividad emocional no solo mejora la interacción, sino que también modula el nivel de confianza que los usuarios depositan en estos sistemas. Se identificaron umbrales en la expresividad emocional donde un exceso de entusiasmo o dramatización puede ser percibido como artificial, generando escepticismo en el usuario.

Además, los resultados refuerzan el marco teórico de la teoría del flujo en interacción humano-ordenador, al sugerir que los asistentes virtuales que logran inducir estados emocionales positivos facilitan experiencias más inmersivas y satisfactorias. De este modo, este estudio contribuye a la literatura sobre interacción humano-ordenador al demostrar cómo la emocionalidad percibida en la voz sintetizada impacta en la confianza y la aceptación de la tecnología.

El cuarto capítulo contribuye al avance de la *Social Response Theory* (SRT), la formación de confianza en la inteligencia artificial y la investigación sobre el boca a boca

electrónico, al demostrar que los asistentes de voz no son meros intermediarios neutrales, sino que moldean activamente el compromiso del usuario en la interacción. Los hallazgos confirman que los usuarios aplican normas sociales humanas a las interacciones con la IA, reforzando la premisa de SRT (Nass & Moon, 2000), pero refinándola al mostrar que la percepción de seguridad, la confianza y la conexión emocional juegan un papel más determinante en la relación usuario-asistente que la mera expresividad vocal.

Uno de los aportes más relevantes es la demostración de que la formación de confianza en asistentes de voz es un proceso secuencial: los usuarios primero deben percibir al asistente como seguro antes de desarrollar confianza, lo que a su vez influye en la intención de compra. Además, este estudio enfatiza que la confianza en las voces sintéticas no es unidimensional, sino que combina evaluaciones racionales (competencia, fiabilidad) con percepciones emocionales. Estos hallazgos refinan las teorías tradicionales sobre la confianza en IA, destacando que la credibilidad de un asistente no solo depende de su precisión, sino también de su capacidad para generar resonancia emocional con el usuario.

Asimismo, este estudio extiende la literatura sobre influencia social y persuasión digital, demostrando que el tipo de voz de un asistente puede moderar el impacto del boca a boca electrónico en la toma de decisiones de compra. Se confirma que la credibilidad de la IA varía en función de su capacidad para generar confianza y conexión emocional, alineándose con las teorías que consideran la IA como un agente de persuasión social (Keller & Fay, 2012).

En términos generales, esta investigación reformula la comprensión del impacto de los asistentes de voz en la toma de decisiones del consumidor, destacando que la influencia de las voces sintéticas no se basa únicamente en atributos funcionales, sino en mecanismos psicológicos y emocionales que median la relación entre el usuario y la tecnología. Con ello, se refinan las teorías existentes sobre la interacción humano-ordenador, la formación de confianza y la persuasión digital, proporcionando un marco más completo para comprender cómo las características de la voz afectan el comportamiento del consumidor en el comercio por voz.

5.3. Implicaciones empresariales

Esta tesis ofrece valiosas implicaciones prácticas para el desarrollo de asistentes de

voz en el comercio electrónico. En línea con las prioridades del Marketing Science Institute (MSI), los estudios incluidos en esta investigación (Capítulos 1, 2, 3 y 4) abordan cuestiones clave sobre la experiencia del usuario en interacciones con tecnología basada en voz. A continuación, se presentan las respuestas a las principales preguntas de investigación, vinculándolas con los capítulos de la tesis.

- *¿Cuál es la mejor manera de medir la experiencia del cliente?*

El capítulo 1 destaca que la experiencia del cliente en comercio por voz no debe medirse solo en términos de usabilidad, sino también en función de percepción de valor, confianza y seguridad en la interacción. El capítulo 2 refuerza esta idea al demostrar que la humanización de la voz influye en la percepción de seguridad y confianza del usuario. Por lo tanto, la medición de la experiencia del cliente debe integrar factores emocionales y cognitivos, permitiendo evaluar de manera más precisa la aceptación y efectividad de estos asistentes en la compra digital.

- *¿Cómo influyen distintos factores en la adopción de nuevas tecnologías en el comportamiento del consumidor?*

El capítulo 1 analiza cómo la adopción del comercio por voz depende no solo de la funcionalidad del asistente, sino también de la experiencia del usuario en la interacción. Se destaca que la confianza y percepción de utilidad son factores clave que determinan la aceptación de esta tecnología, y que el diseño del asistente debe enfocarse en mejorar estos aspectos para incrementar su uso en compras en línea. De forma complementaria, el capítulo 3 demuestra que la entonación y expresividad vocal influyen en la percepción de fiabilidad del asistente, lo que sugiere que la forma en que se presenta la información impacta en la confianza del usuario. Estos hallazgos tienen implicaciones directas en el diseño de asistentes virtuales, indicando que las empresas deben optimizar tanto el contenido informativo como la modulación vocal para generar interacciones más persuasivas y efectivas en el comercio digital.

- *¿Cómo puede el uso de la tecnología mejorar o perjudicar la experiencia del cliente, equilibrando el deseo de conexión humana auténtica con la necesidad de un servicio inmediato y sin fricciones?*

Los resultados del capítulo 2 indican que los asistentes de voz pueden mejorar la

experiencia del usuario cuando ofrecen una sensación de humanización y calidez, generando confianza en la interacción. Sin embargo, si el asistente es percibido como demasiado mecánico, los usuarios pueden sentirse incómodos y evitar su uso. El capítulo 3 refuerza esta idea al mostrar que las voces capaces de transmitir emociones básicas —como la de Siri— generan una mayor conexión emocional y confianza en comparación con otras más neutras o planas. Este equilibrio entre expresividad y naturalidad vocal permite satisfacer la necesidad del usuario de una interacción más humana, sin renunciar a la inmediatez propia de estos sistemas. En el capítulo 4, se analiza cómo la percepción de seguridad y confiabilidad modera la intención de compra en asistentes de voz. Se demuestra que la elección del tipo de voz es un factor crítico en la creación de una experiencia equilibrada entre eficiencia y conexión emocional.

- *¿Cómo pueden optimizarse las técnicas de comunicación persuasiva para aumentar la respuesta del consumidor a los mensajes de marketing?*

La investigación realizada en el capítulo 4 confirma que la percepción de humanidad y confianza en los asistentes de voz influye directamente en su disposición a utilizar estos sistemas para llevar a cabo compras. Se ha observado que los asistentes con una voz más cálida y expresiva generan una mayor sensación de seguridad y confianza, lo que incrementa la intención de compra. Estos resultados sugieren que las estrategias de diseño vocal deben centrarse en optimizar la naturalidad y cercanía de la voz para aumentar la efectividad de los asistentes en contextos de comercio digital.

- *¿Cómo responden los consumidores a diferentes tipos de contenido y formatos en las interacciones en línea?*

El capítulo 3 demuestra que la forma en que se transmite la información influye en la confianza del usuario y en la percepción de humanidad del asistente. Se confirma que las voces con entonación cálida y expresiva generan mayor credibilidad, incluso cuando el contenido informativo es idéntico. Además, el estudio analiza cómo la valencia del mensaje afecta la experiencia del usuario: los mensajes con tono positivo incrementan la respuesta emocional y refuerzan la confianza en el asistente, mientras que los mensajes negativos reducen la respuesta emocional y afectan negativamente a la percepción de confianza. Estos resultados evidencian que tanto el contenido como la forma de presentarlo son clave para generar interacciones eficaces con asistentes virtuales.

- *¿Cómo podemos alinear mejor las estrategias de marketing y comunicación con las preferencias del consumidor?*

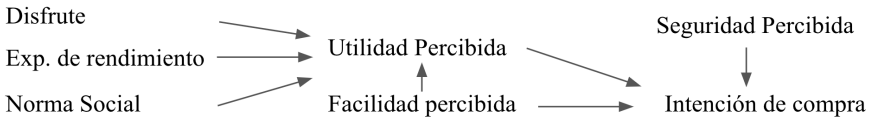
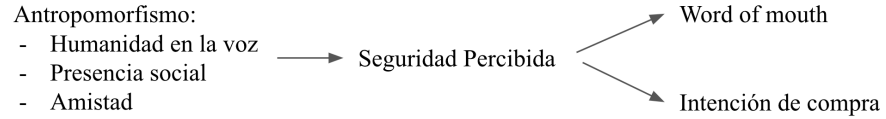
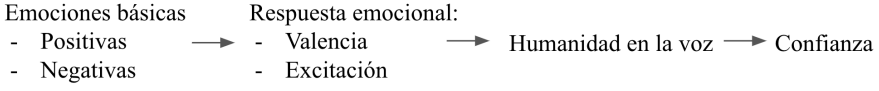
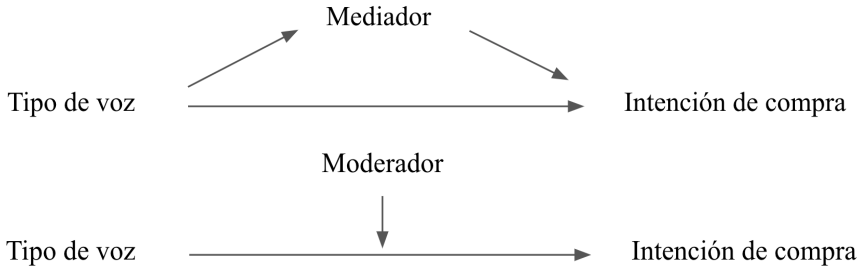
El análisis realizado en el capítulo 4 muestra que el diseño vocal de los asistentes virtuales influye en la intención de compra, no de forma directa, sino a través de mecanismos psicológicos como la percepción de seguridad, la confianza, la atracción social. la humanidad percibida y la respuesta emocional. Se observa que las voces que transmiten mayor calidez y confianza incrementan la disposición del usuario a utilizar el asistente como canal de compra. En consecuencia, las estrategias de marketing deberían centrarse en optimizar las características vocales que fomentan estos procesos, reforzando la percepción de seguridad y facilitando una experiencia de compra más satisfactoria.

Esta tesis ha demostrado que la experiencia del usuario en el comercio por voz no puede evaluarse únicamente en función de la usabilidad o la eficiencia, sino que debe considerar factores psicológicos y emocionales, como la confianza, la seguridad percibida y la humanización de la voz. A lo largo de los cuatro estudios, se ha evidenciado que la percepción de humanidad en los asistentes virtuales es un factor clave en su adopción, influyendo en la disposición del usuario a utilizar estos sistemas como canal para realizar compras.

En particular, se ha observado que la humanización y el tipo de voz impactan directamente en la credibilidad del asistente virtual. Mientras que las voces neutras o robóticas pueden generar escepticismo, aquellas más cálidas y naturales fomentan la confianza del usuario y aumentan su intención de utilizar el asistente como canal de compra. Entre los asistentes analizados, Siri se posiciona como el más efectivo en la construcción de confianza y seguridad percibida. Su voz genera una mayor conexión emocional con el usuario, lo que favorece la aceptación del comercio por voz.

Estos hallazgos proporcionan una base sólida para la optimización del diseño de asistentes de voz en entornos comerciales. Las empresas y desarrolladores deben priorizar la implementación de asistentes con características vocales que refuercen la naturalidad y calidez, asegurando una interacción más persuasiva y efectiva. En definitiva, la mejora de estos sistemas permitirá fortalecer la relación usuario-tecnología y fomentar la adopción masiva del comercio por voz en el futuro. La Tabla 5.1 resume las implicaciones teóricas y prácticas de cada estudio.

Tabla 5.1 Resumen de las principales conclusiones

Marco conceptual	Conclusiones e implicaciones teóricas y empresariales
	**Capítulo 1** - La aceptación de AV depende significativamente de la utilidad percibida, expectativa de rendimiento y seguridad percibida. - El disfrute y la expectativa de rendimiento impactan positivamente en la utilidad, aumentando la intención de usar AV. - La aceptación social no influye en la intención de uso de AV.
	**Capítulo 2** - El antropomorfismo aumenta significativamente con la introducción de la voz en los asistentes independientemente del tipo de voz. - Las variables relacionadas con la humanización incrementan la percepción de seguridad. - La percepción de seguridad influye positivamente en el WOM y la intención de compra.
	**Capítulo 3** - La voz de los AV es capaz de transmitir emociones básicas positivas y negativas. - Las emociones positivas transmitidas por los AV producen una respuesta emocional positiva y por el contrario, las negativas producen una respuesta emocional negativa. - Las emociones transmitidas por la voz producen una respuesta emocional en el usuario que influye directa y significativamente en la percepción de humanidad y en la confianza. - Existen diferencias significativas derivadas del tipo de voz. - Siri destaca como la voz más efectiva en evocar respuestas emocionales positivas. - La valencia del mensaje modera la relación entre respuesta emocional y percepción del AV, siendo los mensajes positivos más efectivos.
	**Capítulo 4** - El tipo de voz del AV influye indirectamente en la intención de compra a través de mediadores (percepción de humanidad, seguridad, confianza, respuestas emocionales y atracción social). - Las voces percibidas como más humanas (especialmente Siri) generan mayores respuestas emocionales positivas, percepción de seguridad, confianza y atracción social, aumentando significativamente la intención de compra de los usuarios. - La intención inicial de compra y el WOM activo están moderados por el tipo de voz del asistente, siendo Siri y Google Assistant más efectivas que Alexa.

5.4. Limitaciones y futuras líneas

Si bien esta tesis contribuye al conocimiento sobre la humanización de los asistentes de voz en el comercio electrónico, presenta ciertas limitaciones que deben considerarse en futuras investigaciones. A continuación, se exponen las principales limitaciones y algunas posibles líneas de estudio para ampliar los hallazgos obtenidos.

Una primera limitación reside en el enfoque teórico adoptado para explicar la adopción del comercio por voz. Aunque se han utilizado modelos consolidados como el TAM (Davis, 1989) y la UGT (Katz et al., 1973), estos priorizan dimensiones racionales y utilitarias, dejando en segundo plano aspectos más emocionales o subjetivos. Además, al tratarse de un estudio transversal, no se ha analizado cómo la confianza o la percepción de seguridad evolucionan con la experiencia acumulada del usuario. Por ello, futuras investigaciones podrían adoptar un enfoque longitudinal, incorporando metodologías cualitativas (como entrevistas o estudios fenomenológicos) para explorar en mayor profundidad las motivaciones, barreras y percepciones individuales en la interacción con los asistentes de voz.

Otra limitación relevante está relacionada con el contexto del experimento, ya que todos los estudios han sido realizados con voces en español de España. Si bien los resultados han mostrado el impacto del diseño vocal en la percepción de humanidad, confianza y seguridad, no es posible generalizar estos efectos a otros idiomas o culturas. Las diferencias prosódicas, fonéticas y culturales pueden influir en la interpretación emocional de la voz (Zhang & Yang, 2022). Por tanto, futuras investigaciones deberían replicar los estudios en distintos idiomas, variantes dialectales y contextos sociolingüísticos, lo cual permitiría comprobar la validez transcultural de los modelos propuestos.

Además, aunque se ha comprobado que la voz del asistente influye en la confianza, la percepción de humanidad y la intención de compra, no se ha considerado el tipo de producto o servicio en juego. Estudios anteriores han demostrado que el nivel de implicación del consumidor y la naturaleza del producto (hedónica vs. utilitaria) influyen en la toma de decisiones y en la percepción del riesgo (Park & Moon, 2003). En este sentido, sería interesante analizar si determinadas características vocales resultan más efectivas dependiendo del tipo de producto.

Asimismo, aunque esta tesis ha analizado la disposición del usuario a utilizar

asistentes de voz para la compra, no se ha abordado directamente la eficacia de las recomendaciones emitidas por estos sistemas. Sin embargo, los resultados sugieren que una voz que genera confianza y conexión emocional —como la de Siri— podría no solo facilitar la adopción del asistente como canal de compra, sino también aumentar la probabilidad de que sus recomendaciones sean aceptadas. Esta hipótesis plantea una prometedora línea futura de investigación sobre la persuasividad de los mensajes comerciales transmitidos por asistentes con distintos estilos vocales.

Por otra parte, sería interesante complementar las técnicas de autoinforme utilizadas en esta tesis con métodos provenientes del neuromarketing, que permitan observar las reacciones fisiológicas del usuario ante distintos tipos de voz. Medidas como la conductancia de la piel, la variabilidad del ritmo cardíaco, el seguimiento ocular o la actividad cerebral podrían ofrecer una visión más objetiva y profunda del impacto emocional de la voz, facilitando la identificación de patrones de respuesta subconsciente ante la interacción con asistentes virtuales.

Por último, sería pertinente explorar diferencias generacionales en la interacción con asistentes virtuales. La familiaridad tecnológica, las expectativas comunicativas y la disposición a confiar en sistemas automatizados pueden variar según la edad (Bawack et al., 2024). Estudiar cómo distintas cohortes generacionales perciben y adoptan estas tecnologías permitiría diseñar experiencias más inclusivas y segmentadas.

En resumen, aunque esta tesis ofrece una visión integral sobre el papel de la voz en la aceptación de los asistentes virtuales para la compra online, queda margen para seguir investigando la interacción entre usuarios y tecnología. Futuros estudios podrían avanzar hacia un diseño vocal más personalizado, adaptado a las características del consumidor y al tipo de producto, e incorporar herramientas neurocientíficas o fisiológicas para analizar con mayor precisión la respuesta emocional en entornos de comercio conversacional.

5.5. Referencias

Bawack, R. E., Bonhoure, E., & Mallek, S. (2024). Why would consumers risk taking purchase recommendations from voice assistants? *Information Technology & People*.

Chen, Q., & Park, H. (2021). How anthropomorphism affects trust in intelligent personal assistants. *Ind. Manag. Data Syst.*, *121*, 2722-2737. https://doi.org/10.1108/imds-12-2020-0761

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13*(3), 319-340. https://doi.org/10.2307/249008

Deshpande, A., Rajpurohit, T., Narasimhan, K., & Kalyan, A. (2023). Anthropomorphization of AI: opportunities and risks. *arXiv preprint arXiv:2305.14784*.

Fakhimi, A., Garry, T., & Biggemann, S. (2023). The Effects of Anthropomorphised Virtual Conversational Assistants on Consumer Engagement and Trust During Service Encounters. *Australasian Marketing Journal*, *31*, 314-324. https://doi.org/10.1177/14413582231181140

Huh, J., Whang, C., & Kim, H.-Y. (2023). Building trust with voice assistants for apparel shopping: The effects of social role and user autonomy. *Journal of Global Fashion Marketing*, *14*(1), 5-19.

Katz, E., Blumler, J. G., & Gurevitch, M. (1973). Uses and Gratifications Research. *The Public Opinion Quarterly*, *37*(4), 509-523.

Keller, E. B., & Fay, B. (2012). *The face-to-face book: Why real relationships rule in a digital marketplace*. Simon and Schuster.

Mari, A., Mandelli, A., & Algesheimer, R. (2024). Empathic voice assistants: Enhancing consumer responses in voice commerce. *Journal of Business Research*, *175*, 114566.

Mori, M. (1970). The uncanny valley. *Energy*, *7*(4), 33-35.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers.

Journal of social issues, *56*(1), 81-103.

Park, C.-W., & Moon, B.-J. (2003). The relationship between product involvement and product knowledge: Moderating roles of product type and product knowledge type. *Psychology & Marketing*, *20*(11), 977-997.

Sun, Y., & Liang, Q. (2023). Research on the Relationship between Self-AI Connection and Customer engagement–Based on the Analysis of Customer AI Trust and AI Anthropomorphism. *Proceedings of the 2023 6th International Conference on Information Management and Management Science*, 48-53.

Zhang, A., & Yang, Q. (2022). To be human-like or machine-like? An empirical research on user trust in AI applications in service industry. *2022 8th International Conference on Automation, Robotics and Applications (ICARA)*, 9-15.