Escuela de Comunicaciones y Periodismo

Universidad Adolfo Ibáñez

Viña del Mar, Chile

Efectos de la transparencia visual en imágenes generadas por inteligencia artificial sobre la percepción de credibilidad, autenticidad y confianza en influencers: un estudio cuasi experimental

Autores:

Catalina Urzúa Galleguillos

Matías Vargas Marambio

Profesora guía:

Dra. Carmina Rodríguez Hidalgo

Tesis presentada para optar al grado de Magíster en Diseño Estratégico de Comunicaciones

30 de abril de 2026

Índice

Resumen

Esta investigación examina el efecto de la transparencia visual —operacionalizada mediante la inclusión de un ícono de advertencia "Información de IA" de Instagram— sobre la percepción de credibilidad, autenticidad y confianza en usuarios chilenos jóvenes frente a publicaciones de influencers en redes sociales. Mediante un diseño cuasi experimental factorial 2×2 (tipo de imagen: influencer humano vs. generado por IA; presencia de ícono: con vs. sin), con una muestra final de N = 201 participantes de entre 18 y 28 años, se aplicó un Análisis Multivariado de Varianza (MANOVA). Los resultados no mostraron diferencias estadísticamente significativas para ninguno de los tres efectos evaluados (ícono, tipo de imagen e interacción), con valores de Pillai's Trace menores a .03 en todos los casos. Los resultados descriptivos, sugieren un patrón de interacción en la variable confianza: en los grupos de influencer generado por IA, la presencia del ícono estuvo asociada a una reducción de 0,72 puntos respecto a la condición sin ícono, mientras que en los grupos de influencer humano la diferencia fue de apenas 0,08 puntos. Estos hallazgos sugieren que el ícono de transparencia, en su formato actual de baja prominencia visual, no logra activar un procesamiento consciente capaz de modificar las evaluaciones de la audiencia. Se discuten las implicancias para el diseño regulatorio de la transparencia en inteligencia artificial y para futuras investigaciones en comunicación digital estratégica.

Palabras clave: Transparencia en IA, influencers digitales, credibilidad percibida, autenticidad percibida, confianza en el contenido, estudio experimental, redes sociales.

Abstract

This study examines the effect of visual transparency—operationalized through the inclusion of an Instagram's "AI Information" warning icon—on perceptions of credibility, authenticity, and trust among young Chilean users when exposed to influencer posts on social media. Using a 2×2 quasi-experimental factorial design (image type: human influencer vs. AI-generated; icon presence: with vs. without), with a final sample of N = 201 participants aged 18 to 28, a Multivariate Analysis of Variance (MANOVA) was conducted. The results revealed no statistically significant differences for any of the three effects assessed (icon, image type, or interaction), with Pillai's Trace values below .03 in all cases. Descriptive results suggest an interaction pattern in the trust variable: within the AI-generated influencer condition, the presence of the icon was associated with a decrease of 0.72 points compared to the no-icon condition, whereas in the human influencer condition the difference was only 0.08 points. These findings suggest that the transparency icon, in its current low-visual-salience format, fails to activate conscious processing capable of modifying audience evaluations. Implications for the regulatory design of transparency in artificial intelligence and for future research in strategic digital communication are discussed.

Keywords: AI transparency, digital influencers, perceived credibility, perceived authenticity, content trust, experimental study, social media.

Agradecimientos

A mi mamá, Angelina, por su comprensión infinita, su amor incondicional y por ser siempre un refugio y un impulso en los momentos en que más lo necesito.

A mi papá, Adrián, por enseñarme con el ejemplo el valor de la responsabilidad, la disciplina y el esfuerzo, y por darme las herramientas para enfrentar cada desafío con carácter y constancia.

A mi hermana, Javiera, por acompañarme de cerca en este camino, por escucharme siempre y por estar presente en cada decisión importante.

A Igna, Clau, May, Jo, Fer, Cami y Nico, por su compañía durante estos años y por una amistad que es, sin duda, lo más valioso que me deja la universidad.

A mi profesora guía, Carmina Rodríguez, por creer en este proyecto desde el inicio y por guiarnos con paciencia y criterio en todo el proceso.

Gracias a mi partner de principio a fin, Matías, por embarcarse en este desafío conmigo, por su actitud positiva inagotable y por estar ahí cuando todo se hacía cuesta arriba.

Y a mí, por sostenerme en los momentos de duda, por seguir adelante incluso cuando no tenía ganas. Porque llegar hasta aquí también es reconocer todo lo que fui capaz de crecer en el proceso.

Catalina Urzúa Galleguillos

A mi papá, Rubén, por abrirme las puertas de la empresa familiar y permitirme dedicar tiempo a este proyecto incluso en horario de trabajo. Su generosidad hizo posible que esta tesis tuviera el espacio que necesitaba.

A mi mamá, Marcela, y a mi polola, Paula, por empujarme a seguir adelante en los momentos en que las ganas simplemente no aparecían. Su apoyo constante fue el impulso que muchas veces necesité para no detenerme.

A mi hermana, Josefina, por estar siempre presente y preocuparse por mí con una dedicación que nunca pasa inadvertida.

A mis amigas del quinto año, Clau, Igna, May y Jo: la universidad en Santiago no habría sido la misma sin ustedes. Gracias por hacer de esta etapa algo que vale la pena recordar.

A mi profesora guía, Carmina, por exigirnos siempre dar lo mejor de nosotros y por enseñarnos, con una paciencia y precisión admirables, a cuidar cada detalle de este trabajo. Su guía fue fundamental para llegar hasta aquí.

Y el agradecimiento más especial va para mi amiga y compañera de tesis, Catalina Urzúa. Juntos enfrentamos esta odisea: las horas interminables escribiendo, pensando y reflexionando; los momentos en que ya no quedaban ganas ni energía, pero aun así encontrábamos la manera de seguir. Esta tesis no habría sido posible sin esa complicidad.

Por último, me agradezco a mí mismo. Este proceso me demostró que soy capaz de conseguir todo lo que me propongo, que la disciplina se construye día a día y que los límites, muchas veces, solo existen en la mente. Como dijo Napoleon Hill: *"Si puedes creerlo, puedes lograrlo."*

Agradezco profundamente a cada persona que formó parte de este camino.

Los quiero mucho a todos.

Matías Vargas Marambio

1. Introducción

Durante las últimas dos décadas, la expansión de las redes sociales ha transformado profundamente las dinámicas de producción, circulación y validación de la información. En este complejo y dinámico contexto, han surgido los denominados influencers digitales, creadores de contenido que poseen una red significativa de seguidores, publican de manera constante en plataformas de redes sociales (como Instagram, TikTok y YouTube) manteniendo de esa forma una relación sostenida y frecuente en el tiempo con sus audiencias. Gracias a dicha relación, los influencers digitales son capaces de afectar decisiones de consumo, opiniones o percepciones en su audiencia (Freberg, Graham, McGaughey & Freberg, 2011).

A partir de la década de 2010, estos actores comenzaron a consolidarse como figuras mediáticas relevantes dentro del ecosistema digital, especialmente en el marco de la denominada economía de la atención, en la que captar y mantener la atención de las audiencias se vuelve un recurso escaso y valioso, lo cual les permite transformar esa visibilidad en valor para actores comerciales y marcas (Arriagada & Ibáñez, 2020).

A diferencia de las celebridades tradicionales, los influencers construyen su influencia mediante la generación permanente de contenido, el desarrollo de una marca personal coherente y la comunicación de experiencias cotidianas que resultan cercanas para sus seguidores. Este proceso favorece la creación de vínculos de confianza y autoridad simbólica, lo cual incrementa su capacidad persuasiva y la credibilidad de los mensajes que difunden (Villegas, 2022; Bautista & Chávez, 2021). En consecuencia, los influencers no solo entretienen, sino que también moldean percepciones, actitudes y opiniones sociales o políticas entre sus audiencias.

La relevancia de estos actores es particularmente significativa entre jóvenes, quienes utilizan las redes sociales como principal fuente de información, desplazando progresivamente a los medios tradicionales (Newman et al., 2023). En el caso chileno, estudios recientes muestran que las personas entre 18 y 29 años privilegian plataformas digitales para informarse, considerando que allí los contenidos resultan más claros y oportunos que en la prensa escrita o la televisión abierta (Feedback Research & Universidad Diego Portales, 2024). De este modo,

los influencers se han transformado en mediadores informativos relevantes dentro de la esfera pública digital.

El poder comunicacional de estos creadores no depende únicamente del tamaño de su audiencia, sino de la relación simbólica que establecen con sus seguidores (De Veirman, Cauberghe & Hudders, 2017; Jin, Muqaddam & Ryu, 2019). La literatura describe este vínculo como una relación parasocial: un lazo unilateral en el que el espectador percibe cercanía, familiaridad e incluso intimidad con una figura pública con la que no interactúa directamente (Horton & Wohl, 1956; Tukachinsky, 2011). Esta percepción de proximidad favorece la atribución de credibilidad, confiabilidad y autenticidad al emisor, fortaleciendo la credibilidad del mensaje que transmite (Sokolova & Kefi, 2020; Lou & Yuan, 2019).

En este escenario, la evaluación de la información en redes sociales depende crecientemente de la percepción de la fuente más que de la verificación directa del contenido (Chaiken, 1980; Metzger, Flanagin & Medders, 2010). Cuando el comunicador es considerado creíble, auténtico y confiable, la evidencia científica indica que sus mensajes tienden a ser aceptados con menor cuestionamiento, operando la credibilidad como un filtro interpretativo de la información.

Sobre este nuevo y complejo ecosistema comunicacional, se incorpora un nuevo elemento: la aparición de influencers generados mediante inteligencia artificial. El desarrollo reciente de herramientas de IA generativa ha permitido crear entidades digitales hiperrealistas, capaces de simular comportamientos humanos, publicar contenido e interactuar con audiencias en redes sociales (Thomas & Fowler, 2021). Estos agentes pueden presentar gestos, voz y personalidad en extremo auténticos comparados con personas reales, lo cual podría dificultar distinguirlos de emisores humanos (Appel, Müller, & Stiglbauer, 2020). Sin duda, esta evidencia nos plantea nuevas interrogantes sobre la evaluación de la fuente y la confianza en los mensajes difundidos por influencers generados por IA.

Diversas investigaciones han demostrado que estos agentes pueden ser percibidos como más creíbles cuando presentan rasgos antropomorfizados y conductas alineadas con códigos sociales humanos, como por ejemplo, atribuciones de capacidades cognitivas, emociones o valores morales similares a los humanos, las cuales influyen positivamente en la credibilidad

percibida y en la relación parasocial con las audiencias (Dabiran et al., 2024). Sin embargo, cabe preguntarse si es que la interpretación del mensaje podría variar, dependiendo de la conciencia del espectador respecto de la naturaleza del emisor, en este caso, de ser un emisor generado por IA. Es decir, el darse cuenta que el comunicador no es humano podría modificar la confiabilidad, credibilidad y autenticidad depositada en él y alterar la evaluación de su autenticidad.

En consecuencia, el problema central ya no radica únicamente en si un influencer artificial puede ser creíble, sino en cómo los usuarios interpretan la información cuando conocen el origen artificial del contenido. Este punto resulta particularmente relevante en un entorno digital donde la producción de imágenes mediante inteligencia artificial (IA) va en aumento, es cada vez más accesible y por ende, la distinción entre lo humano y lo artificial se vuelve progresivamente difusa.

Considerando este contexto del ecosistema digital contemporáneo, la incorporación de inteligencia artificial (IA) generativa en la producción de contenidos visuales en redes sociales ha tensionado los criterios tradicionales de evaluación de la autenticidad, credibilidad y confiabilidad de los mensajes que los usuarios consumen en redes sociales. Diversos estudios han evidenciado que la creciente sofisticación de los sistemas de generación de imágenes dificulta la capacidad de los usuarios para distinguir entre contenidos producidos por personas reales y aquellos generados algorítmicamente, afectando los procesos de evaluación de la fuente y del mensaje (Ohanian, 1990; Willemsen et al., 2024; Lee et al., 2024).

Es relevante entonces que, debido a la emergencia de imágenes hiperrealistas generadas por inteligencia artificial, el fenómeno ha generado crecientes suspicacias en los entornos digitales. Diversos estudios advierten que este tipo de contenidos incrementa el riesgo de desinformación y manipulación persuasiva, al facilitar la creación de representaciones visuales verosímiles que pueden ser utilizadas para inducir interpretaciones erróneas o engañosas sobre personas, hechos o contextos (Chesney & Citron, 2019; Vaccari & Chadwick, 2020). Esta vulnerabilidad no solo afecta a los usuarios expuestos directamente al contenido falso, sino que produce un efecto de segundo orden igualmente preocupante: la proliferación de imágenes artificiales erosiona progresivamente la confianza general en los entornos

digitales, debilitando la capacidad de los usuarios para evaluar la autenticidad de cualquier contenido —incluso el legítimo— y afectando la credibilidad de emisores que no han recurrido a ninguna manipulación (Fletcher et al., 2019; Van der Meer & Jin, 2020).

En el contexto latinoamericano —y particularmente en Chile— este fenómeno adquiere especial relevancia, por un lado, considerando el alto nivel de penetración de redes sociales: en Chile, por ejemplo, el uso de redes sociales alcanza aproximadamente 81,7% de la población, lo que favorece la exposición masiva a contenidos visuales y aumenta la probabilidad de circulación de piezas descontextualizadas o manipuladas en plataformas como Instagram y TikTok, especialmente en ámbitos vinculados a la comunicación política, el marketing digital y la influencia en línea (DataReportal, 2026). Por otro lado, el estudio "Los chilenos y la inteligencia artificial" (Activa Research, 2024) entrega antecedentes cuantitativos relevantes para contextualizar la creciente incorporación de esta tecnología en la vida cotidiana. A nivel de conocimiento, solo un 29,6% de los chilenos declara tener mucho conocimiento sobre cómo la IA interactúa en su vida diaria, mientras que un 38,1% reconoce tener poco o ningún conocimiento, brecha que se profundiza en los grupos socioeconómicos más bajos (44,4% en el segmento D-E). En términos de uso, las personas se relacionan con la IA en promedio 3,7 veces por semana y un 43,3% lo hace todos los días, evidenciando su integración en prácticas cotidianas. Respecto a su valoración, un 47,8% considera que la IA aporta mucho o bastante a su vida y un 56,7% percibe un impacto positivo en la sociedad, percepción que aumenta a 66,6% entre jóvenes de 18 a 30 años. Sin embargo, estos niveles de uso y valoración conviven con una confianza moderada: solo un 29,2% declara confiar en la IA, mientras que un 43,3% manifiesta una confianza media y un 27,5% expresa poca o ninguna confianza. En este contexto, marcado por una alta exposición a contenidos generados por IA, pero niveles intermedios de conocimiento y confianza, resulta pertinente examinar cómo los usuarios evalúan la credibilidad de contenidos digitales mediados por inteligencia artificial, especialmente en entornos como redes sociales, donde estos agentes predominan crecientemente en la oferta de contenidos.

La emergencia de influencers en redes sociales generados por IA posee especial relevancia en contextos donde la imagen cumple una función persuasiva central, como el marketing y la comunicación digital estratégica. Esto especialmente considerando que la literatura reciente

ha encontrado que los usuarios atribuyen niveles significativos de credibilidad y autenticidad a estos agentes, en particular cuando estos presentan altos grados de antropomorfismo —es decir, la atribución de rasgos humanos a entidades no humanas— y que además exhiben coherencia narrativa, entendida como la consistencia entre la apariencia visual del agente, el mensaje que comunica y el contexto discursivo en el que se inserta, lo que favorece la percepción de verosimilitud y naturalidad del contenido (Willemsen et al., 2024; Lee et al., 2024). A pesar de estos efectos encontrados en muestras internacionales, el fenómeno permanece aún poco explorado empíricamente en Chile, en particular los posibles efectos en la percepción de los usuarios de la transparencia visual explícita, entendida como la advertencia directa al usuario de que una imagen ha sido generada mediante inteligencia artificial.

En el contexto chileno, no existen actualmente legislaciones específicas ni estándares obligatorios vigentes que regulen de manera explícita la señalización visual del uso de IA en imágenes difundidas a través de redes sociales. Si bien existe una moción parlamentaria presentada en abril de 2023 orientada a establecer obligaciones de identificación y etiquetado de contenidos generados mediante inteligencia artificial (Aedo et al., 2023), esta fue posteriormente refundida con el proyecto gubernamental Boletín N° 16821-19 y actualmente se encuentra en segundo trámite constitucional en el Senado. Ninguno de los dos instrumentos ha sido promulgado ni implementado, por lo que no constituyen aún un marco normativo aplicable en la práctica.

A diferencia de la identificación de contenidos publicitarios, que cuenta con marcos éticos y legales consolidados, como la obligación de informar de manera clara y verificable cuando un contenido tiene fines publicitarios, conforme a la Ley N.º 19.496 sobre Protección de los Derechos de los Consumidores y a los lineamientos del Consejo de Autorregulación y Ética Publicitaria (CONAR), la identificación explícita de contenido generado por inteligencia artificial queda, en la práctica, a criterio de las plataformas digitales y/o de los propios emisores del mensaje, e incluso sujeta a la interpretación del espectador.

A nivel internacional, la Unión Europea ha avanzado de forma más concreta en esta materia mediante el Reglamento (UE) 2024/1689, conocido como Reglamento de Inteligencia

Artificial (Parlamento Europeo y Consejo de la Unión Europea, 2024), el cual establece obligaciones de transparencia para contenidos generados o manipulados artificialmente que puedan inducir a error al público, como los denominados contenidos sintéticos profundos —en inglés, *deepfakes*—, que son representaciones visuales, sonoras o audiovisuales generadas o alteradas digitalmente mediante inteligencia artificial de manera que simulan de forma convincente a personas, hechos o contextos reales que no ocurrieron o fueron distorsionados. El reglamento, en su artículo 50, exige que los usuarios sean informados de que el contenido ha sido generado o modificado mediante IA, aunque no impone un formato único de identificación visual, permitiendo distintas formas de divulgación, como etiquetas, avisos o descripciones visibles. Estas disposiciones comenzarán a ser plenamente aplicables a partir del 2 de agosto de 2026.

En ausencia de una regulación legal uniforme, las plataformas digitales han implementado de manera autónoma mecanismos de identificación de contenido generado por IA. Redes sociales como TikTok, YouTube e Instagram (Meta) han incorporado políticas que exigen a los creadores declarar cuando un contenido ha sido generado o alterado significativamente mediante IA, especialmente cuando se trata de imágenes, videos o audios realistas. En concreto, TikTok requiere que los creadores etiqueten el contenido generado por IA cuando incluya representaciones realistas de personas, escenas o eventos, aplicando una etiqueta visible que indica que el contenido fue 'divulgado por el creador como generado por IA'; adicionalmente, la plataforma puede aplicar esta etiqueta de manera automática cuando detecta contenido sintético mediante el estándar técnico de Content Credentials (TikTok, 2024). Por su parte, Meta anunció en abril de 2024 e implementó a partir de mayo del mismo año un sistema de etiquetas —inicialmente denominadas 'Hecho mediante IA' y posteriormente renombradas como 'Información de IA'— para informar a los usuarios cuando el contenido fue creado o modificado con inteligencia artificial, apoyándose en indicadores técnicos de la industria y en la autodeclaración de los creadores (Meta, 2024). De este modo, estas plataformas utilizan etiquetas visibles, avisos contextuales o sistemas de detección automática para informar a los usuarios, constituyendo, hasta la fecha, el principal mecanismo operativo de regulación efectiva en el entorno digital actual, en ausencia de marcos normativos legalmente vinculantes.

En síntesis, la identificación de contenidos generados por inteligencia artificial en redes sociales se encuentra actualmente regulada de manera fragmentada, combinando legislación general, autorregulación ética y las normas privadas de las plataformas. En el caso chileno, al momento de escritura de esta tesis, la exigencia de identificación se articula principalmente a través de criterios de transparencia y no engaño, más que mediante la imposición formal legal de un ícono que advierta sobre el contenido generado por de IA, situación que podría evolucionar, a medida que avancen las discusiones legislativas y se consoliden estándares internacionales.

En este escenario, el uso de íconos, etiquetas o advertencias visuales que indiquen "imagen generada por IA" emerge como una práctica incipiente, pero no estandarizada, cuyos efectos sobre la percepción del público no han sido suficientemente estudiados. Desde esta brecha, la presente investigación propone analizar en qué medida la presencia de un ícono de transparencia en la imagen difundida en redes sociales influye en la percepción del receptor, específicamente en variables centrales para la efectividad y ética en la comunicación digital: autenticidad percibida, confianza en el contenido y credibilidad percibida del influencer.

El principal aporte de esta investigación radica en analizar la revelación del origen del contenido (humano vs. generado por IA) como eje central del diseño experimental, entendiendo este elemento como una forma de transparencia visual que actúa como mecanismo comunicacional. Al evaluar el efecto del ícono de IA como estímulo experimental, el estudio contribuye a comprender cómo los usuarios procesan la información sobre el origen del contenido y cómo esta información influye en variables clave para la persuasión digital.

Asimismo, los resultados podrán servir como insumo para discusiones y definiciones muy actuales sobre regulación, diseño ético de plataformas y buenas prácticas en el uso de inteligencia artificial en comunicación estratégica en nuestro país.

Pregunta de investigación

Los antecedentes presentados nos permiten formular la siguiente pregunta de investigación: ¿En qué medida la transparencia en imágenes generadas por IA influye en la percepción de

credibilidad, autenticidad y confianza en usuarios chilenos jóvenes de redes sociales, en comparación con imágenes no generadas por IA?

2.1. Objetivo general

Analizar en qué medida la presencia de un ícono de transparencia en imágenes generadas por inteligencia artificial y el tipo de imagen utilizado —influencer humano vs. generado por IA— influyen en la percepción de credibilidad, autenticidad y confianza percibida de usuarios jóvenes chilenos en redes sociales.

2.2. Objetivos específicos

1- Determinar si la presencia del ícono de transparencia de IA en una publicación de Instagram produce diferencias estadísticamente significativas en los niveles de credibilidad percibida, autenticidad percibida y confianza en el contenido, en comparación con publicaciones que no incluyen dicho ícono.

2- Comparar los niveles de credibilidad percibida, autenticidad percibida y confianza en el contenido entre publicaciones protagonizadas por un influencer humano y publicaciones protagonizadas por un influencer generado por inteligencia artificial.

3- Examinar si existe un efecto de interacción entre el tipo de imagen y la presencia del ícono de transparencia, evaluando si el impacto del ícono sobre las percepciones de los usuarios difiere según el origen humano o artificial de la imagen.

3. Marco teórico

La evaluación de contenidos visuales en redes sociales no responde únicamente a características objetivas del mensaje, sino que involucra una serie de procesos psicológicos y comunicacionales mediante los cuales los usuarios interpretan el origen, la intención y la credibilidad del contenido que consumen. En este marco, el presente estudio organiza su marco teórico en torno a dos grandes bloques conceptuales. El primero agrupa las teorías de emisión del mensaje, es decir, los marcos que permiten explicar los mecanismos subyacentes mediante los cuales el origen, el formato y las señales visuales del contenido influyen en su

procesamiento por parte del receptor. El segundo bloque, desarrollado posteriormente, reúne las teorías de recepción del mensaje, centradas en los constructos perceptuales que permiten operacionalizar los efectos que dicho contenido produce en la audiencia: la credibilidad percibida, la autenticidad percibida y la confianza en el contenido.

Dentro del primer bloque, este estudio se apoya en tres teorías centrales y dos teorías de apoyo. Las teorías centrales son la Teoría de Violación de Expectativas (EVT; Burgoon, 1978), el Modelo Heurístico-Sistemático (HSM; Chaiken, 1987) y la Teoría de Credibilidad de la Fuente (Ohanian, 1990). Estas tres perspectivas son consideradas centrales porque, en conjunto, permiten dar cuenta de los mecanismos más directamente implicados en la respuesta del receptor frente al contenido estudiado: la EVT explica cómo la aparición del ícono de transparencia puede operar como una violación de las expectativas del usuario respecto del origen del contenido; el HSM explica por qué, en un contexto de consumo rápido como las redes sociales, dicha señal puede ser procesada heurísticamente sin activar una evaluación profunda del mensaje; y la Teoría de Credibilidad de la Fuente explica cómo los atributos percibidos del emisor —su experiencia, honestidad y atractivo— median la evaluación global del contenido, independientemente de su origen humano o artificial. Estas tres teorías articulan directamente las hipótesis del estudio y proporcionan el sustento explicativo principal de los efectos esperados.

Las teorías de apoyo son el Modelo de Conocimiento de la Persuasión (PKM; Friestad & Wright, 1994) y la Teoría del Valle Inquietante (Mori, 1970). Aunque no constituyen el núcleo explicativo del diseño experimental, ambas enriquecen la comprensión del fenómeno desde ángulos complementarios: el PKM permite considerar que, en aquellos usuarios que sí detectan e interpretan el ícono como una señal de intento persuasivo, puede activarse un mecanismo de resistencia cognitiva que intensifica la reducción de confianza; mientras que la Teoría del Valle Inquietante ofrece un marco afectivo para comprender por qué la percepción de artificialidad —incluso en ausencia de imperfecciones visuales evidentes— puede generar respuestas emocionales negativas hacia el emisor. A estas perspectivas teóricas se suman los conceptos sustantivos de influencer humano e influencer virtual generado por inteligencia artificial, que constituyen los tipos de emisor analizados en el estudio. A continuación, se detallará cada una de estas perspectivas y su relación con la temática de esta investigación.

3.1. La Teoría de Violación de Expectativas y efecto del ícono de transparencia IA

En primer lugar, la Teoría de la Violación de Expectativas (EVT, por sus siglas en inglés; Burgoon, 1978) explica cómo las personas interpretan y evalúan conductas que se desvían de lo que esperan en una interacción comunicativa. La teoría sostiene que los individuos poseen expectativas previas relativamente estables sobre el comportamiento de otros actores comunicacionales, las cuales se construyen a partir de normas sociales y culturales, características del contexto y del tipo de relación con el emisor. Allí, cuando una conducta se sitúa fuera del rango de comportamientos anticipados, se produce una violación de expectativas que activa procesos de atención, interpretación y evaluación del emisor y de su conducta.

Es importante notar que desde la EVT, las violaciones de expectativas no son inherentemente negativas. La teoría distingue entre violaciones positivas y negativas, dependiendo de la valencia atribuida al comunicador. La valencia de recompensa del comunicador se refiere a la evaluación global que el receptor realiza del emisor como interactuante potencial, considerando atributos como credibilidad, estatus, atractivo o confiabilidad. Así, una misma conducta inesperada puede ser evaluada de manera favorable o desfavorable según quién la realice y el contexto en que ocurre (Burgoon, 1978).

Aplicada al contexto de esta investigación, esta teoría resulta pertinente para comprender un posible efecto del ícono de transparencia en IA como una violación de expectativas comunicacionales en el receptor. En redes sociales, los usuarios suelen desarrollar expectativas implícitas respecto del origen humano de las imágenes publicadas por influencers, particularmente cuando estas se presentan como representaciones cotidianas, personales o auténticas. La aparición de una etiqueta visual que indique que una imagen fue generada mediante inteligencia artificial, puede situarse fuera de ese rango de expectativas, activando un proceso de escepticismo evaluador del contenido y de su emisor.

Debido a ello, el ícono de advertencia IA puede operar como un estímulo inesperado que dirige la atención del receptor hacia el origen artificial del contenido, generando una reevaluación del influencer y del mensaje. De acuerdo con la EVT (Burgoon, 1978), cuando la violación es percibida como negativa, por ejemplo, cuando el emisor es evaluado como menos

confiable o menos auténtico, los resultados comunicacionales podrían ser menos favorables que en situaciones donde las expectativas son confirmadas, en relación a contenido generado y actuado por humanos. Esta concepción permite anticipar que la transparencia visual explícita sobre contenido generado por IA podría reducir percepciones de credibilidad, autenticidad y confianza, especialmente en imágenes de influencers humanos, donde la expectativa de origen no artificial o natural podría ser más chocante para los usuarios.

Considerando lo previamente dicho, la EVT contribuye entonces a sustentar la hipótesis de que la presencia del ícono de transparencia en IA funcionaría como una violación de expectativas que puede derivar en evaluaciones negativas del contenido y del emisor, particularmente cuando la imagen corresponde a un influencer humano.

Una vez producida la activación atencional asociada a una violación de expectativas, resulta necesario comprender cómo los usuarios procesan cognitivamente la información disponible en el mensaje. En ese sentido, el Modelo Heurístico–Sistemático (Heuristic–Systematic Model, HSM; Chaiken, 1987), que detallaremos a continuación, nos explica cómo las personas procesan la información persuasiva a través de dos modalidades cognitivas distintas.

3.2. El Modelo Heurístico-Sistemático y el procesamiento de información persuasiva

El Modelo Heurístico–Sistemático (HSM, por sus siglas en inglés; Chaiken, 1987) explica cómo los receptores emplean distintas formas de procesar la información de un mensaje persuasivo. Éstas dos vías son el procesamiento heurístico y el procesamiento sistemático. El procesamiento heurístico se caracteriza por el uso de “atajos” cognitivos o señales simples que permiten emitir juicios rápidos con bajo esfuerzo mental, mientras que el procesamiento sistemático implica un análisis profundo y detallado del contenido del mensaje, requiriendo así mayor esfuerzo cognitivo y recursos atencionales.

El modelo se articula en torno al principio de suficiencia, según el cual los individuos buscan equilibrar la precisión del juicio con la economía del esfuerzo cognitivo. Las personas procesan la información hasta alcanzar un nivel de confianza considerado suficiente para cumplir sus objetivos, sin invertir más esfuerzo del necesario. Cuando la motivación por la precisión es alta y existen recursos cognitivos disponibles, aumenta la probabilidad de que el

procesamiento sea sistemático, en cambio, cuando la motivación o los recursos son bajos, predomina el procesamiento heurístico.

En el contexto de esta investigación, el ícono de transparencia en IA podría entenderse como una señal heurística que influye en la evaluación del contenido, sin necesidad de un análisis profundo de la imagen. Para muchos usuarios, la mera presencia de un ícono que indica "generado por IA" podría activar juicios rápidos asociados a creencias previas respecto de la artificialidad, de la manipulación o de la falta de autenticidad del contenido, operando como un atajo cognitivo que finalmente reduce la credibilidad y la confianza percibidas.

No obstante, el HSM también permite considerar que, para ciertos usuarios con mayor interés o mayor conocimiento previo sobre inteligencia artificial, la presencia del ícono podría activar un procesamiento más sistemático. En estos casos, el receptor podría interpretar la transparencia bajo un pensamiento masético o informativo, evaluando el contenido de manera más profunda. Sin embargo, el modelo también advierte que ambos tipos de procesamiento pueden coexistir y que las señales heurísticas pueden sesgar el procesamiento sistemático, afectando la evaluación desde un inicio.

Desde esta perspectiva, el HSM (Chaiken, 1987) permite explicar por qué la transparencia visual puede tener efectos negativos predominantes en variables como credibilidad, autenticidad y confianza: en contextos de consumo rápido de contenidos, como redes sociales, podría ser probable que los usuarios recurran a juicios heurísticos basados en señales visuales simples, como el ícono de IA. Esto sustenta la expectativa de que la presencia del ícono reducirá la evaluación positiva del contenido.

3.3. El Modelo de Conocimiento de la Persuasión (PKM)

Además del modo en que se procesa la información, la evaluación del contenido puede verse afectada por la activación de creencias sobre la intención persuasiva del mensaje y de su emisor. Por lo que es relevante adentrarse en la teoría de *Modelo de Conocimiento de la Persuasión* (Persuasion Knowledge Model, PKM); (Friestad & Wright, 1994, PKM, por sus siglas en inglés) que describe el conocimiento que las personas desarrollan acerca de las tácticas, estrategias y motivaciones utilizadas por otros para persuadirlas. El modelo concibe

la persuasión como una interacción dinámica entre un agente de persuasión y un objetivo, en la cual el receptor no es pasivo, sino que utiliza su conocimiento para interpretar y responder estratégicamente a los intentos de influencia.

Para el PKM, lo relevante no es la exactitud del reconocimiento del intento persuasivo, sino la creencia del individuo de que está siendo persuadido. En otras palabras, cuando las personas perciben que un mensaje tiene una intención persuasiva explícita, pueden activar mecanismos de resistencia cognitiva y emocional, lo que tiende a reducir la confianza en el contenido y en el agente que lo emite (Kirmani & Campbell, 2008).

Aplicado a esta investigación, la etiqueta visual que indica que una imagen fue generada por IA puede funcionar como un disparador del conocimiento de la persuasión. La advertencia explícita puede llevar al receptor a reinterpretar el contenido no solo como artificial, sino también como estratégicamente diseñado para influir, especialmente en contextos de marketing de influencers. Este "cambio de significado" transforma una imagen que podría haber sido percibida como espontánea o genuina, en un intento deliberado de persuasión mediado tecnológicamente.

La teoría PKM postula que esta activación de la conciencia persuasiva puede generar desconfianza en el receptor, particularmente cuando el emisor es percibido como un agente externo o impersonal, como ocurre con la inteligencia artificial. La combinación de artificialidad y persuasión explícita puede intensificar el rechazo, reduciendo la credibilidad del contenido y la confianza del espectador. En este sentido, la transparencia no necesariamente aumenta la confianza, sino que puede exacerbar la percepción de manipulación.

Este marco permite sustentar la expectativa de que la presencia del ícono de IA reducirá la confianza y la credibilidad percibidas, especialmente cuando el contenido es interpretado como un intento persuasivo no humano. De este modo, la teoría PKM podría explicar los mecanismos psicológicos que vinculan transparencia, conciencia persuasiva y rechazo del mensaje.

Ahora, más allá de los procesos cognitivos y motivacionales, la literatura en comunicación persuasiva ha destacado también el rol de la fuente en la evaluación de los mensajes, particularmente en contextos de influencia digital. Por ello, la teoría de la fuente y la credibilidad (Ohanian, 1990), desarrollada en el ámbito del marketing y la publicidad, explica cómo las características del emisor influyen en la evaluación del mensaje por parte del receptor.

3.4. La teoría de la fuente y la credibilidad y la evaluación del mensaje

En su teoría de la fuente y de la credibilidad, Ohanian (1990) conceptualiza la credibilidad de la fuente como un constructo compuesto por tres dimensiones: experiencia percibida, confiabilidad (honradez) y atractivo. Estas dimensiones operan conjuntamente en la formación de actitudes hacia el mensaje, la marca y el emisor.

Según este enfoque, los receptores tienden a evaluar los mensajes como más persuasivos y confiables cuando la fuente es percibida como experta, honesta y coherente. La credibilidad no se limita al contenido del mensaje, sino que se construye a partir de la percepción global del emisor y de la congruencia entre su imagen, su rol y el contexto comunicacional.

Diversas investigaciones han examinado cómo la revelación del uso de inteligencia artificial en contenidos digitales puede influir en la forma en que las audiencias evalúan la información que consumen. En particular, el uso de advertencias o etiquetas que indican que un contenido fue generado mediante inteligencia artificial, conocidas como *Etiquetas de divulgación de contenido generado por IA*, introduce señales adicionales que los usuarios consideran al momento de interpretar el origen del contenido.

En esta línea, en un estudio experimental con diseño entre-sujetos, Koning y Voorveld (2025) encontraron que la divulgación del uso de IA en contenidos publicitarios generados por IA puede modificar la confianza que los usuarios depositan en el mensaje y en la organización que lo emite, siendo estos efectos mediados por el conocimiento conceptual sobre IA y el conocimiento persuasivo de los participantes. De manera similar, Morosoli et al. (2025) encontraron, en un estudio experimental con diseño factorial mixto, que los contenidos etiquetados como generados por inteligencia artificial tienden a ser evaluados como menos

creíbles que aquellos que no presentan este tipo de advertencias, aunque el efecto varía según el tipo de contenido y el tema abordado. Asimismo, Schilke y Reimann (2025) advierten, a partir de trece experimentos realizados en distintos contextos profesionales, que la transparencia respecto del uso de IA no produce efectos ambivalentes sino consistentemente negativos sobre la confianza: quienes divulgan el uso de IA son evaluados con menor confianza que quienes no lo hacen, debido a que la divulgación activa percepciones de ilegitimidad sobre el trabajo realizado.

Los estudios anteriores demuestran que la información sobre el origen artificial de un contenido es suficiente para alterar las percepciones de los usuarios. Sin embargo, cabe preguntarse qué ocurre cuando esa información no está disponible de manera explícita: ¿puede la mera apariencia de artificialidad, percibida visualmente sin mediación de etiquetas o avisos, producir respuestas similares? Esta pregunta remite a un fenómeno distinto pero relacionado: la reacción afectiva espontánea ante entidades o representaciones que se perciben como casi humanas, pero no del todo. Para dar cuenta de este proceso, resulta necesario incorporar un marco teórico de naturaleza perceptual y afectiva que complemente los enfoques cognitivos y persuasivos revisados hasta aquí.

3.5. La artificialidad del contenido y la teoría del Valle Inquietante

Finalmente, la quinta perspectiva teórica que integra este marco es la que nos permite abordar las respuestas afectivas asociadas a la percepción de artificialidad, incluso en ausencia de fallas visuales evidentes. Esta es la teoría del Valle Inquietante (Uncanny Valley, Mori, 1970), que describe una relación no lineal entre el grado de similitud con lo humano de una entidad artificial y la afinidad emocional que genera. A medida que aumenta la semejanza humana, la afinidad crece hasta un punto en que una similitud casi humana, pero imperfecta, produce una caída abrupta en la respuesta afectiva, generando sensaciones de extrañeza, inquietud o rechazo.

En desarrollos posteriores, esta teoría ha sido reinterpretada más allá del parecido físico, recalcando más la percepción de artificialidad y la ambigüedad categorial entre lo humano y lo no humano, entendida como la dificultad para clasificar claramente un estímulo dentro de una categoría definida cuando presenta rasgos que lo sitúan simultáneamente entre ambas

(Gutiérrez Aguilar, 2021). Desde esta perspectiva, el rechazo no surge únicamente por la apariencia, sino por la conciencia de que aquello que parece humano no lo es, lo que activa emociones negativas asociadas a la artificialidad.

Aplicada al contexto de esta investigación, la teoría del valle inquietante permite comprender cómo el conocimiento explícito de que una imagen fue generada por IA puede generar menos credibilidad o rechazo, incluso cuando la imagen es visualmente convincente. En este caso, la aversión no se explica por rasgos físicos inquietantes, sino por la identificación consciente del origen artificial del contenido, lo que puede reducir percepciones de calidez, cercanía y autenticidad.

Como hemos visto, en redes sociales, donde los influencers - hasta el momento conformados por humanos en su gran mayoría - suelen construir vínculos simbólicos basados en la cercanía y la identificación, la artificialidad explícita puede interrumpir estos procesos relacionales. Por ello, proponemos que la presencia del ícono de IA podría activar una respuesta emocional negativa, asociada a la percepción de artificialidad, teniendo como posible consecuencia la disminución de la confianza y la credibilidad del contenido, independientemente de su calidad visual.

Las cinco perspectivas teóricas revisadas —el Modelo de Credibilidad de la Fuente, el Modelo Heurístico-Sistemático, la Teoría de Violación de Expectativas, el Modelo del Conocimiento Persuasivo y la Teoría del Valle Inquietante— no operan de manera aislada en este estudio, sino que constituyen el marco interpretativo desde el cual se analizarán los conceptos que se desarrollan a continuación. En conjunto, permiten anticipar de qué manera los usuarios procesan, evalúan y reaccionan afectivamente ante el contenido de influencers en redes sociales, tanto humanos como generados por inteligencia artificial. Los apartados que siguen —referidos al influencer humano, al influencer virtual generado por IA y a las variables de credibilidad, autenticidad y confianza— deben leerse, por tanto, a la luz de estos marcos: cada concepto sustantivo encuentra en ellos su fundamento explicativo y su conexión con las hipótesis del estudio.

3.6. Influencer humano

En el ecosistema digital, un influencer humano puede definirse como una persona real, no necesariamente una celebridad tradicional, que alcanza notoriedad pública al consolidar una audiencia significativa en plataformas digitales (Freberg et al., 2011). Su capacidad de influencia, sin embargo, no depende exclusivamente del volumen de seguidores, sino de su habilidad para sostener una producción constante de contenido, construir una marca personal reconocible y generar una relación emocional y de confianza con su audiencia (Abidin, 2016). En este sentido, los influencers suelen presentarse como "personas comunes", pero con estilos de vida, conocimientos, valores o narrativas identitarias que resultan atractivas para públicos específicos al reforzar su sentimiento de pertenencia a un grupo, favoreciendo la percepción de cercanía y autenticidad, y reforzando su potencial persuasivo (Balmaceda et al., 2022; Villegas, 2022; Guerrero, 2017).

Los influencers humanos producen contenido en formatos múltiples (fotografías, videos breves, transmisiones en vivo, historias efímeras y publicaciones escritas) y abordan temáticas diversas como moda, bienestar, educación financiera, activismo o comunicación política. Esta variedad de formatos y tópicos no solo amplía su alcance, sino que permite modular el tono comunicacional (informativo, recreativo, comercial o ideológico) y adaptar el mensaje a las lógicas de consumo rápido y visual que caracterizan a plataformas como Instagram y TikTok.

A pesar de que la investigación sobre influencers ha sido profusa, principalmente al revisar la literatura podemos notar que esta se ha desarrollado desde enfoques descriptivos o correlacionales. Aunque este tipo de estudios han contribuido a comprender a los influencers como nuevos líderes de opinión en entornos digitales (Freberg et al., 2011) y como actores con capacidad de incidir en audiencias digitales (por ejemplo, evidenciando que atributos como el número de seguidores influyen en la percepción de popularidad, credibilidad y en las actitudes hacia la marca) (De Veirman, Cauberghe & Hudders, 2017), existe aún menor evidencia de carácter experimental que nos permita aislar el efecto de elementos específicos de presentación digital, por ejemplo, su formato visual o el uso de etiquetas, sobre variables perceptivas relevantes que nos permitan medir y comprender su influencia. Estudios previos

han notado cómo las variables de credibilidad, autenticidad y confianza son clave para entender la influencia de estos nuevos actores en el ecosistema digital.

En particular, Freberg et al. (2011) conceptualizan a los social media influencers como “un nuevo tipo de endosante independiente de terceros que moldea las actitudes de la audiencia a través de blogs, tuits y el uso de otras redes sociales”, lo que permite comprenderlos directamente como líderes de opinión en el ecosistema digital. Por su parte, De Veirman et al. (2017) muestran empíricamente que “Los influencers con una gran cantidad de seguidores tienden a ser percibidos como más populares, lo cual, a su vez, mejora las actitudes hacia la marca.”, ejemplificando cómo la literatura ha tendido a centrarse en relaciones entre variables más que en manipulaciones experimentales.

Esta brecha se vuelve especialmente pertinente en el escenario contemporáneo, donde el contenido puede estar mediado por tecnologías que modifican la apariencia del emisor o del mensaje, tales como la inteligencia artificial generativa (Intriago et al., 2024; Balmaceda et al., 2022; Guerrero, 2017).

3.7. Influencer virtual generado por inteligencia artificial

Desde una perspectiva conceptual, la inteligencia artificial generativa se refiere a sistemas capaces de producir contenido nuevo —por ejemplo, imágenes, videos, avatares y elementos audiovisuales— a partir de datos de entrada y patrones aprendidos, mediante técnicas de aprendizaje automático. En lugar de limitarse a replicar información existente, estos modelos aprenden la estructura y regularidades de los datos para generar nuevas instancias plausibles, es decir, contenidos originales que mantienen coherencia con los datos sobre los cuales fueron entrenados (Goodfellow et al., 2014). En el marco del marketing de influencers, estas herramientas permiten diseñar perfiles virtuales, generar piezas visuales y ajustar la presentación del contenido de forma escalable, con el objetivo de maximizar la visibilidad y la interacción en plataformas regidas por lógicas algorítmicas. Sin embargo, aunque estas tecnologías pueden optimizar métricas de interacción, ello no implica necesariamente aumentos equivalentes en credibilidad o confianza, especialmente si los usuarios perciben artificialidad o activan sospechas sobre intención persuasiva (Torronteras et al., 2025).

A modo de contextualización, uno de los primeros trabajos en utilizar explícitamente la denominación AI influencer es el de Thomas y Fowler (2021), quienes analizan el uso de estos agentes como endosantes de marca, incorporando el término dentro del campo del marketing digital. Posteriormente, la literatura ha tendido a conceptualizar a los AI influencers como una subcategoría dentro de los influencers virtuales, diferenciándolos en función del uso de inteligencia artificial para generar contenido o interactuar con audiencias. En esta línea, investigaciones más recientes reconocen que el concepto no constituye una categoría teórica independiente, sino una variante terminológica dentro de un ecosistema conceptual más amplio que incluye también nociones como virtual influencers o CGI influencers (Ju et al., 2024).

Un influencer virtual generado por inteligencia artificial generativa corresponde a una figura digital hiperrealista, creada mediante tecnologías computacionales (por ejemplo, gráficos por computadora y/o modelos de IA generativos) que simulan un actor humano en redes sociales. Esto les permite construir una identidad consistente (nombre, apariencia, estilo comunicacional), una trayectoria narrativa y una comunidad de seguidores (Moustakas et al., 2020; Lou et al., 2023). A diferencia del influencer humano, este tipo de agente no posee vivencia personal ni intencionalidad propia; su desempeño comunicacional depende del diseño y control ejercido por desarrolladores, marcas o agencias. Pese a ello, estos perfiles pueden alcanzar altos niveles de realismo social, especialmente cuando incorporan rasgos antropomorfizados —apariencia y/o comportamiento humano— que permiten a la audiencia interpretarlos como “socialmente válidos” dentro del ecosistema digital (Willemsen et al., 2024). No obstante, la forma hiper realista humana no es la única forma que pueden adoptar los influencers IA. De hecho, esta nueva generación de influencers puede adoptar múltiples formas, que van desde personajes animados no humanos hasta creaciones con apariencia humana que pueden resultar tan realistas que llegan a ser indistinguibles de personas reales (Angmo, Mahajan y da Silva Oliveira, 2024; Kim et al., 2024).

La literatura reciente indica que los influencers virtuales pueden ser evaluados favorablemente en términos de autenticidad percibida, credibilidad percibida y confianza en el contenido bajo ciertas condiciones. Un mecanismo explicativo es la heurística de máquina, según la cual algunos usuarios tienden a atribuir a los sistemas automatizados cualidades de

mayor objetividad o menor sesgo, lo que puede elevar la evaluación del emisor artificial en comparación con emisores humanos, especialmente en individuos con creencias positivas sobre la objetividad tecnológica (Sundar, 2008; Sundar & Kim, 2019; Lee et al., 2024).

Pese a ello, esta perspectiva no es concluyente ni universal. Diversos marcos teóricos permiten inferir un efecto opuesto, en el que los usuarios evalúan más negativamente a las máquinas que a los humanos. En particular, la teoría del valle inquietante sugiere que, cuando un agente artificial alcanza altos niveles de realismo, pero sigue siendo percibido como no humano, puede generar sensaciones de extrañeza o rechazo, afectando negativamente su evaluación (Mori, 1970; Appel, Müller & Stiglbauer, 2020). Asimismo, investigaciones recientes en el ámbito de los influencers virtuales muestran que la revelación del origen artificial del contenido puede disminuir la credibilidad y la confianza percibidas en comparación con emisores humanos (Morosoli et al., 2024; Schilke et al., 2025).

Desde la perspectiva de la credibilidad de la fuente, este efecto puede explicarse porque los agentes artificiales carecen de atributos tradicionalmente asociados a la confiabilidad, como la experiencia vivida, la intencionalidad o la capacidad moral, elementos que inciden en la evaluación del emisor (Ohanian, 1990; Lou & Yuan, 2019). En este sentido, mientras algunos enfoques sugieren una sobrevaloración de lo automatizado, otros evidencian una penalización de lo artificial, configurando un escenario teórico ambivalente. Esta tensión justifica la necesidad de investigar empíricamente cómo los usuarios evalúan a los influencers generados por IA, particularmente en contextos donde su naturaleza artificial es conocida o explícita, dando sustento a la pregunta de este estudio.

En el plano internacional, uno de los casos más emblemáticos de influencer virtual es Lil Miquela (@lilmiquela), quien cuenta con aproximadamente 2,4 millones de seguidores en Instagram y explicita, poniendo “Robot” en su descripción, que se trata de una figura generada mediante inteligencia artificial. Su contenido replica las dinámicas propias de los influencers humanos, combinando publicaciones de estilo de vida, moda y reflexiones personales. A nivel visual, presenta una estética consistente: si bien cambia de vestuario, mantiene rasgos estables como el peinado, lo que refuerza su identidad digital. Sus publicaciones incorporan distintos tipos de encuadre (desde primeros planos hasta planos

generales) con el fin de generar cercanía, destacar el vestuario y situar al personaje en diversos contextos. No obstante, en los planos más cerrados, su rostro puede adquirir un aspecto irreal o ligeramente caricaturesco, lo que acentúa su carácter artificial. Además, en el perfil de Lil Miquela se presencia que ha participado activamente en campañas con marcas globales, posicionándose como un actor relevante dentro del ecosistema del marketing de influencia, lo que evidencia el potencial comercial de este tipo de perfiles (Moustakas et al., 2020; Lou et al., 2023). Su perfil además reproduce formatos característicos del trabajo influencer descritos en la literatura, especialmente la combinación entre autorrepresentación cotidiana, moda, estilo de vida y construcción de marca personal, es decir, contenidos centrados en sostener una identidad reconocible, visualmente coherente y socialmente identificable ante la audiencia (Abidin, 2016; Arriagada & Ibáñez, 2020).

A nivel latinoamericano, y particularmente en Chile, es posible identificar el caso de @joseeefasorel, influencer virtual que actualmente cuenta con 43,9 mil seguidores en Instagram. En su biografía se presenta como “Influencer-AI” y “Digital Bestie”, declarando explícitamente su naturaleza artificial. Su contenido se centra principalmente en publicaciones en torno a lógicas de estilo de vida cotidiano, inspiración estética y simulación de intimidad digital, es decir, publicaciones que buscan construir familiaridad con la audiencia a través de rutinas, reflexiones, registros personales y fragmentos conversacionales, formato muy cercano al que la literatura identifica como central en la economía relacional de los influencers (Abidin, 2016; Freberg et al., 2011), más que en contenido abiertamente publicitario. Por ejemplo, ha compartido publicaciones simulando su presencia en eventos como Lollapalooza Chile 2026, así como extractos de un podcast propio llamado “La Vida Con Josefa”, difundidos en formato reel. A nivel visual, su apariencia alcanza altos niveles de realismo en videos, aunque ciertas expresiones pueden percibirse levemente como rígidas o robóticas. A diferencia de otros perfiles orientados al respaldo comercial, no se observa una presencia evidente de marcas en sus publicaciones, ya que los productos o elementos que aparecen (como alimentos o vestuario) no presentan identificación comercial explícita. Este contraste entre perfiles que sí integran colaboraciones de marca y otros que priorizan contenido cotidiano resulta importante para comprender cómo distintos niveles de intencionalidad comercial podrían influir en la percepción de credibilidad, autenticidad y

confianza, variables centrales en el análisis de la influencia digital (Lou & Yuan, 2019; De Veirman et al., 2017).

En este sentido, la presencia o ausencia de contenido asociado a marcas no solo constituye una diferencia descriptiva entre perfiles, sino un elemento que podría influir en la forma en que los usuarios evalúan a estos emisores. La literatura en marketing de influencia ha demostrado que la identificación de intenciones persuasivas, como ocurre en el contenido patrocinado, puede activar procesos de escepticismo en los receptores, afectando la forma en que interpretan y valoran el mensaje (De Veirman, Cauberghe & Hudders, 2017; Evans et al., 2017). Desde esta perspectiva, el carácter comercial del contenido no es neutro, sino que incide directamente en la construcción de percepciones sobre el emisor.

En particular, la autenticidad percibida, entendida como el grado en que el contenido y su emisor son evaluados como genuinos y no artificialmente construidos, puede verse tensionada cuando los usuarios reconocen una intención promocional explícita (Balmaceda et al., 2022; Villegas, 2022). Del mismo modo, la confianza en el contenido, definida como la disposición del receptor a aceptar la información como fiable, puede disminuir cuando se percibe que el mensaje responde a intereses comerciales (Evans et al., 2017; Li & See-To, 2024). Finalmente, estas evaluaciones inciden en la credibilidad percibida del emisor, entendida como la valoración global de su competencia, honestidad y confiabilidad, la cual se construye a partir de la coherencia entre el contenido, la identidad del influencer y el contexto en que se presenta (Ohanian, 1990).

De este modo, el contraste entre influencers virtuales que integran activamente marcas en su contenido y aquellos que priorizan narrativas cotidianas sin explicitación comercial no solo refleja distintas estrategias de construcción de identidad digital, sino que también permite problematizar cómo la intencionalidad persuasiva puede incidir en variables perceptuales. Este punto resulta relevante para el presente estudio, en la medida en que permite vincular las características del contenido con la evaluación que los usuarios realizan sobre emisores generados mediante inteligencia artificial.

Los apartados anteriores han permitido caracterizar a los dos tipos de emisor que conforman el diseño experimental de este estudio: el influencer humano, cuya capacidad de influencia

descansa en la construcción de vínculos de cercanía, autenticidad y confianza con su audiencia; y el influencer virtual generado por inteligencia artificial, cuya evaluación por parte de los usuarios depende de factores como el grado de realismo visual, la percepción de artificialidad y la presencia o ausencia de señales explícitas sobre su origen. Ambos tipos de emisor son analizados en este estudio a través de tres constructos perceptuales que la literatura ha identificado como centrales para comprender cómo las audiencias evalúan el contenido en entornos digitales: la credibilidad percibida, la autenticidad percibida y la confianza en el contenido. Estos constructos constituyen las variables dependientes del estudio y conforman el segundo bloque del marco teórico, orientado a las teorías de recepción del mensaje.

3.8. Influencia mediada por algoritmos y señales visibles

La influencia en redes sociales se encuentra crecientemente mediada por algoritmos que seleccionan y jerarquizan contenidos según criterios automatizados, condicionando la exposición de los usuarios a determinadas personas, ideas o productos. En este contexto, métricas visibles como “me gusta”, comentarios o reproducciones actúan como señales públicas que pueden reforzar la percepción de legitimidad y autoridad del influencer, operando como validación social y algorítmica, aunque no siempre reflejen influencia real (Arriagada & Valderrama, 2023). Esta mediación resulta relevante para esta tesis, porque el ícono de IA es precisamente, una señal visible que puede competir con otras señales del entorno (estética, métricas, reputación del emisor) y orientar la evaluación del contenido.

3.9. Teorías de recepción del mensaje: credibilidad, autenticidad y confianza percibidas

Si las teorías anteriores explican los mecanismos desde los cuales se emite y procesa el contenido, las variables de recepción permiten operacionalizar los efectos que ese contenido produce en la audiencia. En este estudio se consideran tres constructos perceptuales centrales en la literatura sobre influencers y comunicación digital:

3.9.1. Autenticidad percibida

La autenticidad percibida refiere al grado en que el usuario percibe la imagen y su emisor como genuinos, reales y no artificialmente construidos (Balmaceda et al., 2022; Villegas, 2022). En el contexto del marketing de influencers, la autenticidad no constituye una propiedad objetiva del contenido, sino una evaluación subjetiva construida por el receptor a partir de señales visuales, narrativas y contextuales que le permiten juzgar si lo que observa corresponde a una representación genuina del emisor (Abidin, 2016). Esta percepción es especialmente relevante en plataformas como Instagram, donde los influencers construyen su capital simbólico precisamente sobre la base de parecer cercanos, espontáneos y no mediados por intereses comerciales explícitos (Arriagada & Valderrama, 2023). Sin embargo, dicha autenticidad no es puramente espontánea: se produce dentro de un entorno condicionado por normas de plataforma y expectativas de audiencia, generando tensiones entre lo genuino y lo performativo. Esta percepción es además frágil; cambios percibidos como incoherentes u oportunistas pueden erosionar rápidamente la autenticidad atribuida al influencer, afectando también la credibilidad y la confianza (Intriago et al., 2024).

La literatura ha demostrado que la autenticidad percibida opera como un mediador clave entre la identidad del influencer y su capacidad persuasiva. Audrezet et al. (2020), en un estudio cualitativo basado en entrevistas en profundidad y observación de colaboraciones entre influencers y marcas, muestran que la autenticidad no debe entenderse como un rasgo estable del influencer, sino como un resultado dinámico que emerge de la gestión de tensiones entre intereses personales y exigencias comerciales. Los autores identifican que las colaboraciones con marcas pueden amenazar la percepción de autenticidad cuando los seguidores perciben incongruencia entre el contenido patrocinado y la identidad del influencer, y que los influencers recurren a dos estrategias principales para mantenerla: la autenticidad basada en la pasión —destacando la coherencia entre los productos y sus intereses personales— y la autenticidad basada en la transparencia —reconociendo explícitamente la naturaleza comercial del contenido— (Audrezet et al., 2020). En una línea complementaria, Gohil et al. (2025) muestran que cuando los influencers son percibidos como auténticos, los usuarios tienden a confiar más en ellos y a mostrar mayor disposición a aceptar sus recomendaciones, siendo la confianza el mecanismo que media esta relación. Así, la

autenticidad funciona como una señal que facilita la credibilidad del mensaje y reduce la desconfianza hacia el contenido (Lou & Yuan, 2019; Gohil et al., 2025).

Por el contrario, cuando se activan sospechas sobre la naturaleza construida o artificial del contenido —por ejemplo, ante la presencia de un ícono que indica origen en IA—, la autenticidad percibida tiende a disminuir, afectando la evaluación global del emisor (Morosoli et al., 2025; Schilke & Reimann, 2025). Así como las colaboraciones percibidas como incongruentes afectan negativamente la autenticidad, la revelación del origen artificial del contenido puede operar de manera similar, debilitando esta percepción al introducir una señal que cuestiona la coherencia entre el emisor, su identidad y el contenido que publica (Audrezet et al., 2020). En este sentido, si la transparencia en IA reduce la autenticidad percibida, es esperable que también impacte indirectamente en la confianza y en la evaluación global del mensaje, lo que refuerza la pertinencia de analizar estos efectos de manera conjunta en un diseño experimental como el propuesto (Gohil et al., 2025). En el presente estudio, la autenticidad percibida se entiende como la valoración del usuario respecto de si la imagen y la persona que aparece en ella son genuinas y no el producto de una construcción artificial intencionada.

3.9.2. Confianza en el contenido

La confianza en el contenido refiere al nivel de seguridad y fiabilidad que el receptor atribuye al contenido y a su fuente, expresándose en la disposición a aceptar el mensaje sin cuestionamiento o sospecha activa (Ohanian, 1990; Lou & Yuan, 2019). Desde la perspectiva de la credibilidad de la fuente, esta disposición se construye progresivamente a partir de la percepción de honestidad, competencia y coherencia del emisor: cuando el receptor evalúa favorablemente estos atributos, tiende a aceptar el mensaje como fiable sin necesidad de verificar activamente su veracidad (Ohanian, 1990). En entornos digitales, esta confianza no depende únicamente de la veracidad objetiva de la información, sino de señales perceptuales asociadas al emisor —como su coherencia, transparencia y comportamiento previo— que permiten al usuario inferir si el contenido es digno de ser creído (Li & See-To, 2024).

En el contexto del marketing de influencers, la confianza en el contenido se construye a partir de la relación simbólica que el creador establece con su audiencia, donde la cercanía, la

consistencia comunicacional y la percepción de autenticidad actúan como fundamentos clave para aceptar el mensaje difundido (Lou & Yuan, 2019). Factores como la coherencia entre el estilo de comunicación y los valores proyectados, la transparencia sobre colaboraciones comerciales y la percepción de experiencia genuina en el tema abordado contribuyen igualmente a construir o erosionar esta confianza (Evans et al., 2017; Li & See-To, 2024). En cuanto al rol específico de la transparencia sobre el uso de IA, Li y See-To (2024) precisan que cuando los usuarios cuentan con información clara sobre el origen y el proceso de producción del contenido su disposición a confiar en él puede aumentar, pero cuando esa información activa cuestionamientos sobre la naturalidad, el control o la intencionalidad del mensaje, la confianza disminuye. Investigaciones recientes confirman esta dirección, mostrando que la divulgación del uso de inteligencia artificial introduce incertidumbre sobre la intencionalidad del mensaje y activa sospechas sobre su origen y propósito, afectando negativamente la confianza en el contenido (Koning & Voorveld, 2025; Schilke & Reimann, 2025). En el presente estudio, la confianza en el contenido se entiende como la disposición del receptor a aceptar como fiable la información transmitida por la publicación observada, independientemente de si ha verificado o no su veracidad.

3.9.3. Credibilidad percibida

La credibilidad percibida constituye la evaluación global de la competencia, honestidad y confiabilidad del emisor del mensaje, considerando la coherencia entre su imagen, el contenido que produce y el contexto en que se presenta (Ohanian, 1990). A diferencia de la confianza —que refiere principalmente a la disposición del receptor hacia el mensaje— y de la autenticidad —que evalúa la genuinidad del emisor—, la credibilidad integra ambas dimensiones en un juicio más amplio sobre la fuente como actor comunicacional. Ohanian (1990) la conceptualiza como un constructo tridimensional compuesto por la experiencia percibida del emisor, su confiabilidad y su atractivo, dimensiones que en conjunto determinan cuán persuasivo y creíble resulta el mensaje y que influyen en la disposición del receptor a aceptarlo como válido.

En redes sociales, esta evaluación no se explica únicamente por atributos internos del emisor, sino por dinámicas relacionales: la construcción de comunidad, la interacción sostenida y la

percepción de cercanía con el público contribuyen a reforzar la credibilidad atribuida por la audiencia (Fernández, 2017; Guerrero, 2017). Asimismo, se articula con señales de plataforma —interacciones visibles, consistencia comunicacional, estilo discursivo— y con la percepción ética del emisor, incluyendo expectativas de transparencia y coherencia entre el mensaje y la identidad presentada (Li & See-To, 2024; Arriagada & Valderrama, 2023).

En el ámbito del marketing de influencers, la credibilidad de la fuente ha sido identificada como uno de los predictores más robustos de la efectividad del mensaje: los influencers percibidos como expertos, honestos y coherentes con su identidad digital generan mayor aceptación del contenido y actitudes más favorables hacia las marcas que promueven (Lou & Yuan, 2019; De Veirman et al., 2017). Sin embargo, esta evaluación puede verse alterada cuando se introducen señales que cuestionan la naturaleza del emisor o del mensaje. En particular, una etiqueta explícita que indique que una imagen fue generada por IA puede operar como una señal que reconfigura la interpretación del mensaje y del emisor, reduciendo la credibilidad percibida al interrumpir la coherencia esperada entre el influencer, su historia personal y el contenido que publica, e introduciendo dudas sobre su competencia real, su honestidad y el grado de control sobre el contenido difundido (Morosoli et al., 2025; Koning & Voorveld, 2025). En el presente estudio, la credibilidad percibida se entiende como la valoración que el receptor realiza del emisor en términos de su competencia, honradez y confiabilidad general, a partir de la observación de la imagen y del contenido que esta transmite.

4. Hipótesis

En conjunto, todas estas teorías permiten comprender por qué la transparencia visual en contenidos generados por inteligencia artificial, es decir, advertir al usuario que un contenido ha sido generado por IA, no necesariamente se podría traducir en mayores niveles de confianza, autenticidad y credibilidad, sino que podría activar procesos de violación de expectativas, procesamiento heurístico, conciencia persuasiva, cuestionamiento de la fuente y respuestas afectivas negativas frente a la artificialidad. Estos mecanismos resultan aplicables en el contexto del marketing de influencers, donde la percepción de autenticidad, credibilidad y confianza marcan la diferencia entre la relación simbólica del emisor y de la

audiencia. Por ello, a partir del estado del arte sobre esta materia y de nuestra pregunta de investigación nos es posible formular las siguientes hipótesis de investigación:

- **H1 - Efecto principal de la transparencia IA (ícono presente sí/no):** La presencia del ícono de transparencia IA tendrá un efecto negativo significativo sobre la evaluación del influencer y del contenido, reduciendo de manera conjunta los niveles de credibilidad percibida, autenticidad percibida y confianza en el contenido, en comparación con las condiciones sin ícono IA.

- **H2 (efecto principal):** Las imágenes de influencers generadas por IA serán significativamente menores, en términos de credibilidad percibida, autenticidad percibida y confianza en el contenido, comparadas a las imágenes de influencers humanos.

- **H3 (efecto de interacción):** El efecto negativo del ícono de transparencia IA sobre la evaluación del influencer y del contenido, será significativamente mayor en términos de credibilidad percibida, de autenticidad percibida y de confianza percibida, cuando la imagen corresponda a un influencer humano, que cuando corresponda a un influencer generado por IA.

5. Métodos

5.1. Descripción de la muestra

La muestra original estuvo compuesta por 319 personas. Como se estila en estudios experimentales, basados en la verificación de manipulación, se aplicó un criterio de exclusión: los participantes de los grupos con ícono que indicaron no haberlo visto al responder la pregunta Q19 fueron descartados, siguiendo el procedimiento estándar para este tipo de estudios (Oppenheimer et al., 2009). En estudios experimentales similares con muestras estudiantiles chilenas, Rodríguez Hidalgo (2018) aplicó el mismo tipo de filtro al excluir participantes que seleccionaron incorrectamente el tipo de audiencia o escribieron actualizaciones de estado incoherentes, con el objetivo de asegurar que cada persona hubiera

procesado correctamente la condición experimental asignada. Tras aplicar este criterio, la muestra final quedó compuesta por N = 201 participantes, distribuidos en cuatro grupos experimentales.

En concreto, para llegar a esa cantidad de participantes, antes de analizar los datos, se revisó la base para descartar respuestas que pudieran afectar la validez del estudio. Se eliminaron los participantes de los grupos con ícono (Grupos 1 y 3) que indicaron no haber visto ningún ícono en la imagen, es decir, quienes no se dieron cuenta de que había una etiqueta de transparencia en la foto que les tocó ver. Mantener esos casos habría sido un problema, porque si una persona no vio el ícono, la condición experimental a la que fue asignada no tuvo ningún efecto real sobre ella (Oppenheimer et al., 2009). En total se excluyeron 118 personas: 53 del Grupo 1 y 65 del Grupo 3.

La muestra final quedó compuesta por N = 201 participantes, cuya distribución por condición se puede consultar en la Tabla 1:

Tabla 1

Distribución de participantes por grupo experimental (N = 201)

Grupo	Tipo de imagen	Ícono de transparencia	*n*	%
1	Influencer humano	Con ícono	30	14.9%
2	Influencer humano	Sin ícono	86	42.8%
3	Influencer IA	Con ícono	30	14.9%
4	Influencer IA	Sin ícono	55	27.4%
Total	—	—	201	100%

Nota. Los grupos con ícono (G1 y G3) quedaron con menos participantes que los grupos sin ícono (G2 y G4) por efecto del criterio de exclusión aplicado.

La mayoría de los participantes contaba con educación universitaria completa o en curso (76,0%), mientras que un 13,3% adicional reportó tener estudios de postgrado, lo que eleva a un 89,3% la proporción de participantes con formación universitaria o superior. Dicho esto, los participantes reportaron que usan redes sociales con alta frecuencia a lo largo del día, perfil que se ajusta al de jóvenes adultos chilenos, usuarios activos de plataformas digitales.

5.2. Diseño experimental

Se realizó un estudio con un enfoque cuantitativo de tipo cuasi experimental, con un diseño factorial 2x2, que apuntó a aislar y analizar tanto los efectos principales como los efectos de interacción entre dos variables independientes: el tipo de imagen del influencer (humano vs. generado por IA) y la presencia de un ícono de transparencia (con ícono vs. sin ícono). Esta estructura generó cuatro condiciones experimentales a las que los participantes fueron asignados de manera no aleatoria. Este diseño factorial permite evaluar tanto los efectos principales de cada variable independiente, como su posible efecto de interacción sobre las variables dependientes (Ricks, 2002), lo cual resulta especialmente útil cuando se busca determinar si el impacto de una condición depende del nivel de la otra.

La lógica del diseño experimental factorial ha sido ampliamente utilizada en investigaciones sobre percepciones del consumidor ante estímulos comunicacionales. Por ejemplo, Ricks (2002) empleó un diseño 2×2×2 para examinar cómo distintos tipos de filantropía corporativa afectaban las evaluaciones de marca y las intenciones de compra, utilizando el MANOVA como prueba principal para evaluar efectos simultáneos sobre múltiples variables dependientes correlacionadas.

De manera similar, McGinn (2013) implementó un diseño factorial entre sujetos con doce condiciones para estudiar cómo diferentes estrategias de encuadre en publicidad de servicios influían en la comprensión y actitud hacia el anuncio, midiendo las respuestas del consumidor mediante escalas Likert de siete puntos. Ambos antecedentes respaldan la pertinencia del diseño adoptado en este estudio para el análisis de efectos perceptuales en contextos de comunicación persuasiva.

Condiciones experimentales (4 estímulos)

1. Imagen de influencer real (humano) + ícono “Generado por IA”
2. Imagen de influencer real (humano) + sin ícono IA
3. Imagen de influencer generada por IA + ícono “Generado por IA”
4. Imagen de influencer generada por IA + sin ícono “Generado por IA”

Este diseño permite evaluar tanto el efecto directo de la transparencia visual con respecto a una imagen generada por IA, como las interacciones entre el origen de la imagen y la presencia o ausencia del ícono “Generado por IA”, sobre las variables dependientes.

5.3. Estímulo experimental: Imágenes

Se utilizaron cuatro imágenes como estímulos experimentales, diseñadas para representar las cuatro condiciones del diseño factorial 2×2. El punto de partida fue una imagen extraída de un reel de Instagram de una influencer humana real[1], en la que aparece presentando un protector solar. Se eligió este producto por tratarse de un bien de bajo involucramiento simbólico, es decir, que no activa esquemas identitarios ni aspiracionales en el receptor (Zaichkowsky, 1985), lo que permite focalizar la evaluación en las variables de interés y reducir interferencias persuasivas ajenas al diseño experimental (Ricks, 2002; Petty & Cacioppo, 1986). Adicionalmente, el protector solar presenta alta plausibilidad de aparición en contenidos de influencers en plataformas visuales como Instagram, lo que contribuye a la verosimilitud del estímulo.

A partir de esta imagen base se generaron cuatro versiones. Para las condiciones con ícono, se incorporó editorialmente el ícono de “Información de IA” replicando de manera exacta la posición en que este aparece habitualmente en los reels de Instagram, con el fin de no alterar la composición visual ni generar quiebres de naturalidad. Para las condiciones de influencer de IA, la imagen original fue procesada mediante el sistema de generación de imágenes de ChatGPT, solicitando explícitamente la réplica exacta del estímulo original (prompt consultable en el anexo). Tras varios intentos, se seleccionó la versión con mayor grado de fidelidad visual respecto de la imagen humana. A esta imagen generada por IA se le añadió posteriormente el ícono siguiendo el mismo criterio de posicionamiento, completando así el

[1] Laura Iturrieta. Instagram: @lauriturri

conjunto final de cuatro estímulos: (G1) influencer humana con ícono, (G2) influencer humana sin ícono, (G3) influencer de IA con ícono y (G4) influencer de IA sin ícono.

Con el fin de resguardar derechos de autor y de imagen, se realizaron modificaciones en el nombre de la marca del producto y en la identidad nominal de la influencer, en conformidad con la legislación chilena vigente (Ley N.º 17.336 sobre Propiedad Intelectual). Asimismo, se incorporó texto descriptivo en formato de viñetas con características del producto, coherente con el estilo comunicacional habitual de la influencer en sus contenidos de Instagram y con las prácticas propias de la denominada *visibility labour* (Abidin, 2016). Se replicaron también la tipografía, la paleta cromática y el estilo visual de sus reels, con el objetivo de mantener un alto grado de organicidad en los estímulos. En función de todo lo anterior, se considera que los cuatro estímulos cumplen con los criterios de control experimental, realismo visual y resguardo ético necesarios para su utilización en el presente estudio. A continuación se presentan los cuatro estímulos experimentales utilizados en el estudio, correspondientes a cada una de las condiciones del diseño factorial 2×2:

Figura 1. Imagen de influencer real (humano) + ícono “Información de IA”

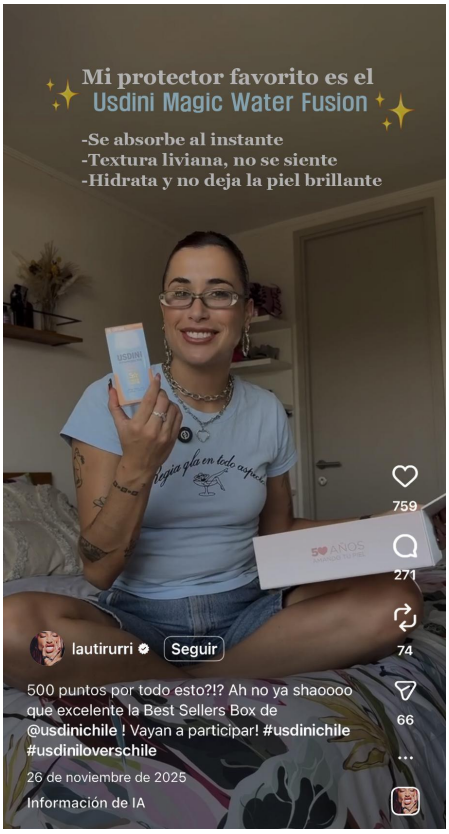


Figura 2. Imagen de influencer real (humano) + sin ícono “Información de IA”

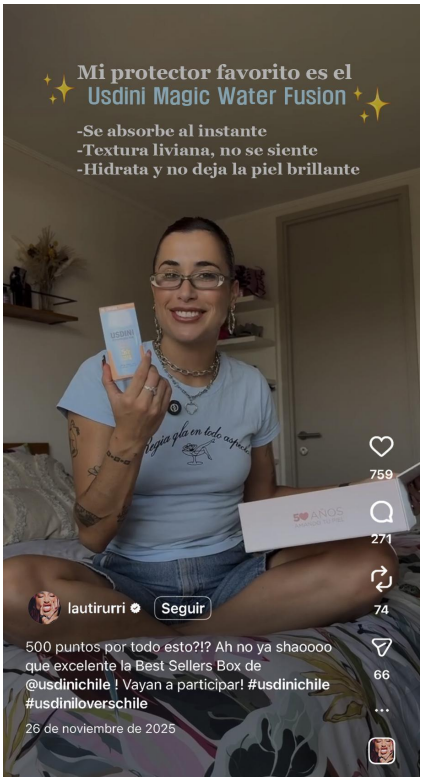


Figura 3. Imagen de influencer generada por IA + ícono “Información de IA”

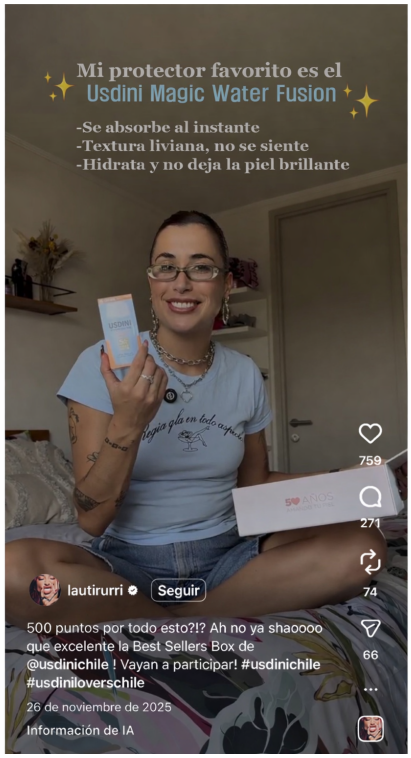
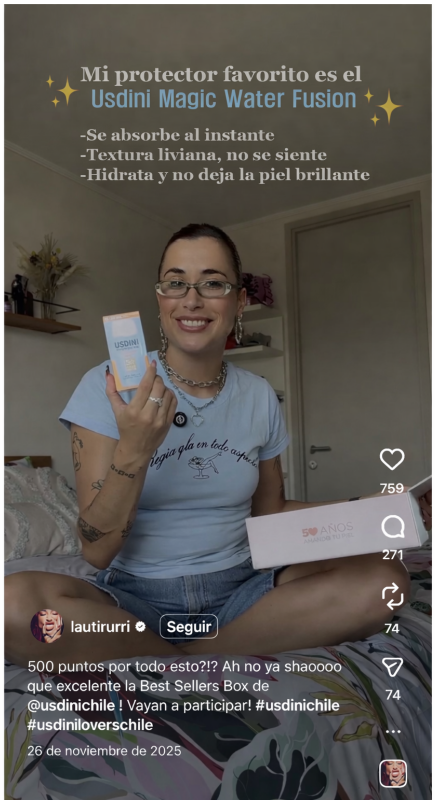


Figura 4. Imagen de influencer generada por IA + sin ícono “Información de IA”



5.4. Variables

El estudio considera dos variables independientes y tres variables dependientes, organizadas en un diseño factorial 2×2.

5.5. Variables independientes

La primera variable independiente es la condición de transparencia, operacionalizada mediante la presencia o ausencia de un ícono visual que indica que la imagen fue generada por inteligencia artificial. Esta variable tiene dos niveles: con ícono y sin ícono.

La segunda variable independiente es el tipo de imagen, operacionalizada mediante el origen de creación del contenido visual. Esta variable también tiene dos niveles: imagen de influencer humano e imagen de influencer generado por inteligencia artificial.

La combinación de ambas variables independientes da lugar a cuatro condiciones experimentales: (G1) influencer humano con ícono, (G2) influencer humano sin ícono, (G3) influencer de IA con ícono y (G4) influencer de IA sin ícono.

5.6. Variables dependientes

Las tres variables dependientes corresponden a constructos perceptuales medidos mediante escalas Likert de 7 puntos (1 = Muy en desacuerdo; 7 = Totalmente de acuerdo), cuyos fundamentos conceptuales fueron desarrollados en el apartado 3.9 del marco teórico.

La credibilidad percibida fue medida a través de cinco ítems (Q1–Q5): (1) "La imagen me parece creíble", (2) "La información que transmite la imagen me parece confiable", (3) "Percibo coherencia entre la imagen y lo que comunica", (4) "La imagen me parece verosímil" y (5) "El contenido de esta imagen no es engañoso".

La autenticidad percibida fue medida mediante cinco ítems (Q6–Q10): (1) "El contenido de la imagen se siente genuino", (2) "La imagen me parece natural", (3) "La persona en la imagen me parece auténtica", (4) "Siento que la imagen representa una situación realista" y (5) "Percibo el contenido de la imagen como fiel a la realidad".

La confianza en el contenido fue medida a través de cinco ítems. Los ítems Q11, Q12, Q13 y Q15 utilizaron una escala Likert de 7 puntos (1 = Muy en desacuerdo; 7 = Totalmente de acuerdo): (1) "Confiaría en una publicación como esta", (2) "El contenido me inspira confianza", (3) "Me sentiría cómodo/a interactuando con este tipo de publicación (por ejemplo, comentar, darle like o compartir)" y (4) "Me fiaría de este contenido al momento de formar una opinión". El ítem Q14 —"¿Qué tan identificado/a se siente con el estilo del contenido observado?"— utilizó una escala Likert de 7 puntos con anclajes distintos (1 = Nada identificado; 7 = Totalmente identificado), por lo que si bien forma parte de la escala de confianza, sus anclajes fueron adaptados para capturar el grado de identificación del receptor con el estilo del contenido. El ítem Q13 —"Me sentiría cómodo interactuando con una cuenta que transmita este tipo de contenido (por ejemplo, comentar, darle like o compartir)"— completa los cinco ítems de esta variable.

5.7. Participantes

La muestra inicial estuvo compuesta por 319 jóvenes chilenos entre 18 y 28 años, rango etario correspondiente a la etapa de adultez emergente (Arnett, 2000). En términos de distribución etaria, el grupo más numeroso correspondió a participantes entre 22 y 25 años (39,3%), seguido por el grupo de 26 a 28 años (36,7%) y el de 18 a 21 años (22,7%). Un 1,3% de los participantes reportó una edad superior al rango definido, por lo que dichos casos fueron excluidos del análisis. En cuanto al género, el 54,0% de los participantes se identificó como hombre, el 45,3% como mujer y un 0,7% seleccionó la opción “Otro / Prefiero no responder”.

Adicionalmente, y en línea con los estándares metodológicos en investigación experimental, se aplicaron criterios de depuración de la muestra asociados al adecuado procesamiento de la manipulación experimental. Específicamente, se excluyeron aquellos participantes pertenecientes a las condiciones con ícono que declararon no haber percibido dicho elemento en el chequeo de manipulación, dado que en estos casos no es posible asegurar la exposición efectiva al estímulo experimental. Este procedimiento es consistente con prácticas ampliamente documentadas en la literatura, que recomiendan descartar casos que no reconocen la condición asignada o que fallan controles de manipulación o atención, con el fin de resguardar la validez interna del diseño.

En total, se excluyeron 118 casos: 53 correspondientes a la condición de influencer humano con ícono y 65 a la condición de influencer generado por IA con ícono. Tras la aplicación de estos criterios, la muestra final quedó conformada por N = 201 participantes, distribuidos en las cuatro condiciones experimentales: influencer humano con ícono (n = 30), influencer humano sin ícono (n = 86), influencer IA con ícono (n = 30) e influencer IA sin ícono (n = 55).

5.8. Instrumento de medición

Las tres variables dependientes —credibilidad percibida, autenticidad percibida y confianza en el contenido— se midieron con escalas de tipo Likert de siete puntos (1 = Muy en desacuerdo, 7 = Totalmente de acuerdo), formato ampliamente utilizado en la literatura sobre evaluación de comunicación persuasiva (McGinn, 2013; Ricks, 2002). Cada escala

estuvo compuesta por entre cuatro y cinco ítems diseñados para capturar distintas dimensiones de cada constructo.

Verificación de manipulación

El cuestionario también incluyó preguntas de verificación de manipulación orientadas a comprobar si los participantes percibieron correctamente el tipo de imagen (Q16, Q17, Q18) y si recordaron haber visto el ícono (Q19). Este tipo de verificaciones es un componente metodológico habitual en diseños experimentales con estímulos visuales, ya que permite asegurarse de que los participantes efectivamente procesaron la condición a la que fueron expuestos (Rodríguez Hidalgo, 2018; Oppenheimer et al., 2009).

5.9. Procedimiento

El cuestionario fue administrado en línea mediante la plataforma Google Forms. Al ingresar al formulario en línea, a cada participante les fue mostrado el consentimiento informado. Cada participante fue dirigido a una única condición experimental según la opción seleccionada al inicio del formulario, donde se le asignó una de las cuatro versiones de la imagen estímulo. Se les pidió observar la imagen durante aproximadamente 15 segundos antes de responder el cuestionario, para simular la experiencia de ver contenido en redes sociales. El diseño siguió un esquema experimental ampliamente utilizado en estudios sobre comunicación y redes sociales, que contempla la exposición a estímulos visuales estandarizados antes de la medición de las variables dependientes (véase, por ejemplo, Rodríguez Hidalgo, 2018), con el propósito de controlar las condiciones de exposición y facilitar la comparación entre grupos. El texto completo del experimento puede encontrarse en la sección apéndice.

1. Los participantes accedieron al cuestionario en línea mediante la plataforma Google Forms, donde se les presentó el consentimiento informado, el cual debían aceptar para continuar con el estudio.
2. Cada participante fue asignado a una única condición experimental, correspondiente a una de las cuatro versiones del estímulo visual. Luego, se les presentó la imagen y se les solicitó observarla durante aproximadamente 15 segundos, con el objetivo de simular una exposición breve a contenido visual en redes sociales.

3. Los participantes respondieron las escalas correspondientes a las variables dependientes del estudio: credibilidad percibida, autenticidad percibida y confianza en el contenido, medidas mediante ítems tipo Likert de siete puntos.
4. Se incorporaron preguntas de chequeo de manipulación, orientadas a verificar si los participantes habían identificado correctamente tanto el tipo de imagen presentada como la presencia o ausencia del ícono de transparencia.

5. Finalmente, los participantes respondieron preguntas de carácter sociodemográfico.

6. Los participantes recibieron una sesión de *debriefing* o retroalimentación al finalizar el experimento, en la que se explicó brevemente el objetivo real del estudio.

El procedimiento se organizó en una secuencia estructurada propia de diseños experimentales en ciencias sociales, que incluyó consentimiento informado, recolección de datos iniciales, exposición controlada a un estímulo, medición de variables dependientes y verificación de la manipulación experimental, con el objetivo de garantizar la estandarización de la experiencia y la validez de las mediciones. El procedimiento se organizó en una secuencia estructurada propia de diseños experimentales en ciencias sociales, que incluyó consentimiento informado, recolección de datos iniciales, exposición controlada a un estímulo, medición de variables dependientes y verificación de la manipulación experimental, con el objetivo de garantizar la estandarización de la experiencia y la validez de las mediciones (Shadish et al., 2002).

6. Resultados y análisis de datos

6.1. Fiabilidad de las escalas

Las tres variables del estudio —credibilidad, autenticidad y confianza— se midieron con grupos de preguntas que en conjunto forman una escala. Para verificar si cada grupo de preguntas mide efectivamente aquello que busca evaluar, se calculó el coeficiente Alpha de Cronbach (Cronbach, 1951), que indica qué tan consistentes son las respuestas entre sí dentro de cada escala. Se expresa en porcentaje: mientras más alto, más coherente es la escala. Valores sobre el 80% se consideran buenos (Nunnally & Bernstein, 1994).

Los resultados fueron los siguientes: la escala de credibilidad explicó el 81.7% de variación consistente entre sus ítems; la de autenticidad, el 86.6%; y la de confianza, también el 86.6%. Las tres escalas superan el umbral del 80%, lo que confirma que las preguntas de cada bloque miden de forma coherente el mismo concepto. Esto permite utilizarlas con mayor confianza y solidez en los análisis posteriores.

El ítem Q14 fue excluido de este análisis porque usaba un formato de respuesta distinto al resto del cuestionario, lo que lo hacía incompatible con las demás preguntas.

6.2. Análisis Factorial Confirmatorio

Además de comprobar la fiabilidad, se realizó un Análisis Factorial Confirmatorio (AFC) para verificar que la estructura del cuestionario efectivamente correspondiera a los tres factores planteados en el diseño del estudio. En términos metodológicos, este análisis permite determinar si los ítems agrupados en cada escala presentan consistencia y coherencia interna, de acuerdo con la estructura teórica previamente establecida (Brown, 2015; Kline, 2016).

El modelo evaluado agrupó los 14 ítems (Q1 a Q13 y Q15) en tres factores correspondientes a credibilidad, autenticidad y confianza, permitiendo que los tres factores estuvieran relacionados entre sí, lo cual tiene sentido, dado que son conceptos vinculados (Ohanian, 1990; Lou & Yuan, 2019). Para evaluar si el modelo se ajusta bien a los datos, se usaron cuatro índices de ajuste estándar en la literatura: CFI y TLI (valores sobre .90 indican buen ajuste) y RMSEA y SRMR (valores bajo .08 indican buen ajuste) (Hu & Bentler, 1999; MacCallum et al., 1996).

Los altos porcentajes de consistencia interna obtenidos en todas las escalas, junto con las elevadas correlaciones entre cada pregunta y su bloque respectivo, son señales de que la estructura de tres factores encuentra respaldo en los datos. La existencia de una relación entre los tres factores también justifica su análisis conjunto mediante MANOVA, en lugar de abordarlos de manera independiente: cuando las variables dependientes están relacionadas, analizarlas en conjunto entrega una imagen más completa y evita cometer más errores estadísticos de los necesarios (Tabachnick & Fidell, 2013).

6.3. Estadística descriptiva por grupo

La Tabla 2 muestra los promedios de cada grupo en las tres variables.

Tabla 2

Medias y desviaciones estándar de credibilidad, autenticidad y confianza por grupo

Grupo / Condición	*n*	Credibilidad M (DE)	Autenticidad M (DE)	Confianza M (DE)
G1: Humano + Con ícono	30	4.34 (1.27)	3.83 (1.56)	3.23 (1.42)
G2: Humano + Sin ícono	86	4.38 (1.02)	3.96 (1.36)	3.31 (1.36)
G3: IA + Con ícono	30	4.06 (1.43)	4.11 (1.77)	2.98 (1.64)
G4: IA + Sin ícono	55	4.39 (1.01)	4.23 (1.21)	3.70 (1.37)
Media general	201	4.32 (1.14)	4.02 (1.42)	3.33 (1.43)

Nota. M = promedio; DE = desviación estándar. Escala: 1 = Muy en desacuerdo, 7 = Totalmente de acuerdo.

Al analizar los promedios, destaca la diferencia en confianza dentro de los grupos de influencer IA: los participantes que vieron la imagen con ícono (G3) obtuvieron un promedio de 2.98, mientras que quienes la vieron sin ícono (G4) llegaron a 3.70, una diferencia de 0.72 puntos. En cambio, entre los grupos de influencer humano (G1 y G2), la diferencia es casi inexistente (3.23 vs. 3.31). Mientras, en las variables de credibilidad y autenticidad percibidas, los cuatro grupos muestran puntajes bastante semejantes entre sí.

Verificación de supuestos

Antes del análisis principal, se verificó que los datos cumplieran las condiciones necesarias para usar el MANOVA. Se comprobó si los datos de cada grupo seguían una distribución normal y si las varianzas eran similares entre grupos.

Para la normalidad, se usó la prueba de Shapiro-Wilk en los grupos pequeños (G1 y G3, con n = 30) y Kolmogorov-Smirnov en los más grandes (G2 y G4). La mayoría de los grupos y variables cumplieron el supuesto. Las excepciones fueron la variable confianza en G1 y G3, y

la credibilidad en G3, donde la distribución se alejó de la normalidad. Para la homogeneidad de varianzas (prueba de Levene), credibilidad y autenticidad percibidas mostraron varianzas distintas entre grupos, mientras que confianza sí cumplió ese supuesto.

Estas desviaciones no representan una limitación significativa para el análisis. El MANOVA tolera bien estas situaciones cuando las muestras son de tamaño razonable (Tabachnick & Fidell, 2013), y el estadístico Pillai's Trace —que es el que se reporta en los resultados— es precisamente el más robusto frente a este tipo de irregularidades (Field, 2018).

6.4. MANOVA Factorial 2×2

Para poner a prueba las tres hipótesis del estudio se realizó un Análisis Multivariado de Varianza (MANOVA) factorial 2×2. Este análisis evalúa si el tipo de imagen (influencer humano vs. influencer IA) y la presencia del ícono (con vs. sin) produjeron diferencias en las tres variables dependientes —credibilidad, autenticidad y confianza— analizadas en conjunto. Analizar las tres variables a la vez, en lugar de por separado, es lo correcto cuando están relacionadas entre sí, porque hacerlo por separado aumenta la posibilidad de encontrar diferencias significativas simplemente por azar (Tabachnick & Fidell, 2013; Hair et al., 2019) y de encontrar errores del tipo I (Field, 2018).

Tabla 3

Resultados del MANOVA: efectos principales e interacción

¿Qué se evaluó?	Pillai's Trace	*F* aproximado	*p*	Significativo	Efecto
Tipo de imagen (humano vs. IA)	.028	F(3, 195) = 1.89	.132	No	Mínimo
Presencia del ícono (con vs. sin)	.018	F(3, 195) = 1.21	.309	No	Mínimo
Interacción entre ambos factores	.022	F(3, 195) = 1.43	.235	No	Mínimo

Nota. N = 201; gl error = 195. Se reporta Pillai's Trace por ser el estadístico más robusto del MANOVA ante las condiciones de esta muestra (Field, 2018). Un resultado es significativo cuando $p < .05$.

El MANOVA no encontró diferencias estadísticamente significativas para ninguno de los tres efectos evaluados. Ni el tipo de imagen, ni la presencia del ícono, ni la combinación de ambos factores lograron producir diferencias relevantes en cómo los participantes percibieron la credibilidad, autenticidad y confianza del contenido. Los valores de Pillai's Trace son muy pequeños en los tres casos, lo que indica que los efectos, de existir, serían de magnitud mínima (Cohen, 1992).

Para identificar en cuál de las tres variables se concentran las tendencias observadas, se examinaron también los efectos de forma individual por variable. Estos análisis son solo exploratorios —no cambian la conclusión principal del MANOVA— pero ayudan a entender el patrón de los datos. La confianza fue la única variable que se acercó a la significación en el factor ícono, $F(1, 197) = 2.64$, $p = .106$, aunque sin llegar al umbral requerido de $p < .05$. Las otras variables —credibilidad y autenticidad— mostraron valores de p muy alejados de la significación en todas las fuentes de variación.

6.5. Análisis de efectos de interacción

La Figura 1 permite ver visualmente si el efecto del ícono fue distinto dependiendo del tipo de imagen. Cada gráfico muestra una variable dependiente: en el eje horizontal está el tipo de imagen (humano vs. IA); las dos líneas representan las condiciones con y sin ícono; y el eje vertical muestra el promedio de cada grupo. Las barras de error representan ±1 error estándar de la media.

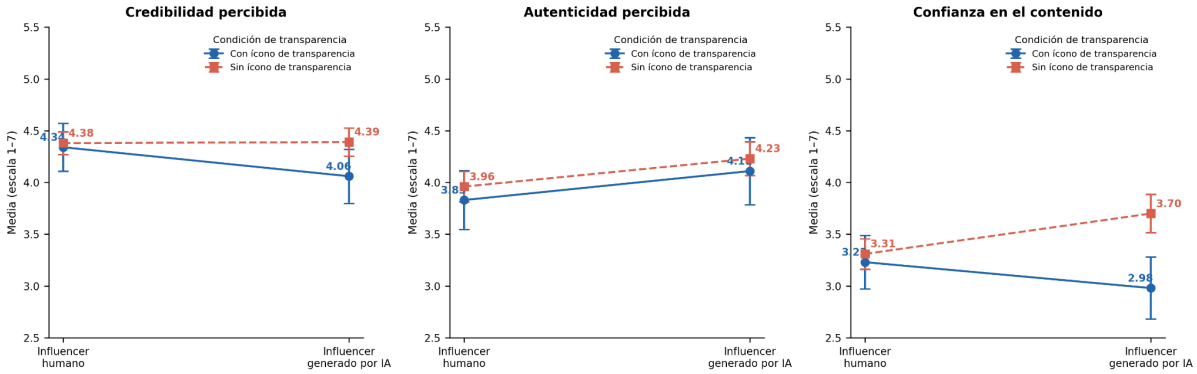


Figura 1. Efectos de interacción: Tipo de imagen × Presencia de ícono de transparencia IA. G1: Humano + Con ícono (n = 30); G2: Humano + Sin ícono (n = 86); G3: IA + Con ícono (n = 30); G4: IA + Sin ícono (n = 55). Las barras de error representan ±1 error estándar de la media.

En los gráficos de credibilidad y autenticidad, las dos líneas son prácticamente paralelas. Eso significa que el ícono tuvo un efecto muy parecido —y pequeño— sin importar si la imagen era de un influencer humano o de IA. El gráfico de confianza, sin embargo, muestra algo distinto: en los grupos con influencer humano, el ícono casi no marca diferencia (G1: M = 3.23 vs. G2: M = 3.31). Pero en los grupos con influencer de IA, la diferencia es bastante mayor: quienes vieron la imagen con ícono reportaron una confianza notablemente más baja que quienes la vieron sin él (G3: M = 2.98 vs. G4: M = 3.70; diferencia de 0.72 puntos). Las líneas se separan visiblemente en la condición de IA, lo que es el patrón típico de una interacción.

Dicho esto, este efecto de interacción no alcanzó la significancia estadística en el MANOVA (p = .235), por lo que no puede afirmarse con certeza que esta diferencia sea real y no producto del azar. En cuanto a H3, el patrón encontrado va en la dirección contraria a lo que la hipótesis preveía: el ícono reduce la confianza más en la condición de IA que en la humana, y no al revés.

Tabla 4

6.6. Síntesis de resultados por hipótesis

	Hipótesis	Resultado MANOVA	Decisión	¿Qué encontramos?
H1	El ícono reducirá la credibilidad, autenticidad y confianza en comparación con los grupos sin ícono	Pillai = .018, F(3, 195) = 1.21, p = .309	No confirmada	El análisis no encontró diferencias significativas entre tener o no el ícono. La confianza mostró una tendencia descriptiva, pero no llega al nivel de significación requerido.
H2	Los influencers IA obtendrán menores puntajes de credibilidad, autenticidad y confianza que los humanos	Pillai = .028, F(3, 195) = 1.89, p = .132	No confirmada	No se observaron diferencias estadísticamente significativas según el tipo de imagen. Asimismo, en los grupos sin ícono, el influencer generado por inteligencia artificial presentó puntuaciones similares e incluso levemente superiores al influencer humano en las dimensiones de autenticidad y confianza.

H3	El efecto negativo del ícono será mayor en el influencer humano que en el de IA	Pillai = .022, F(3, 195) = 1.43, p = .235	No confirmada	La interacción no fue significativa. El patrón descriptivo en confianza es al revés de lo esperado: el ícono tiene mayor impacto cuando la imagen es de IA (diferencia de 0.72 puntos) que cuando es humana (diferencia de 0.08 puntos).

Nota. Las tres hipótesis se evalúan en base al estadístico multivariado del MANOVA como prueba principal.

7. Discusión

El presente estudio tuvo como objetivo analizar en qué medida la presencia de un ícono de transparencia en imágenes generadas por inteligencia artificial y el tipo de imagen utilizado —influencer humano versus influencer generado por IA— influyen en la percepción de credibilidad, autenticidad y confianza de usuarios jóvenes chilenos en redes sociales. Para ello, se examinó si el ícono de transparencia producía diferencias significativas en dichas percepciones, si el tipo de imagen (influencer humano vs. generado por IA) generaba distintos niveles de credibilidad, autenticidad y confianza, y si ambos factores interactuaban entre sí de manera significativa.

Los resultados de este estudio no apoyaron ninguna de las tres hipótesis planteadas. El MANOVA factorial 2×2 no arrojó diferencias estadísticamente significativas ni para el tipo de imagen (influencer humano vs. generado por IA), ni para la presencia del ícono de transparencia, ni para la interacción entre ambos factores. A primera vista, esto podría interpretarse como un resultado fallido o poco informativo. Sin embargo, la literatura metodológica postula que los efectos nulos son informativos cuando se producen bajo condiciones bien controladas y con instrumentos válidos (Cohen, 1992): en ese caso, su

ausencia comunica algo sobre el fenómeno estudiado, no sobre las deficiencias del diseño. En esta investigación, los tres valores de Pillai's Trace resultaron con valores menores a .03, lo cual indica que, de existir efectos reales, su magnitud sería muy pequeña. La discusión que sigue articula y ordena qué significan nuestros hallazgos hipótesis por hipótesis, pone en diálogo a los resultados con la literatura citada en el marco teórico y entrega sugerencias y aprendizajes concretos para futuras investigaciones.

H1: El ícono de transparencia no redujo la credibilidad, autenticidad ni confianza

La primera hipótesis predecía que la presencia del ícono de “Información de IA” reduciría los niveles de credibilidad percibida, autenticidad percibida y confianza en el contenido, en comparación con las condiciones sin ícono. Esta hipótesis no fue confirmada dado que el análisis multivariado no encontró diferencias significativas asociadas a la presencia o ausencia del ícono. Una primera explicación posible para este resultado es que el ícono de transparencia, tal como está diseñado e implementado actualmente en Instagram, carece de la carga comunicacional suficiente para alterar percepciones complejas de multiconstructos, como la credibilidad o la autenticidad. Este hallazgo es coherente con lo señalado por Koning y Voorveld (2025), quienes encontraron que los efectos de las advertencias de contenido elaborado por IA sobre la confianza de los usuarios son altamente sensibles al diseño, la visibilidad y el contexto de la etiqueta. Cuando la divulgación pasa inadvertida o no activa asociaciones cognitivas claras, su impacto se diluye. El propio sistema de etiquetado de Meta fue objeto de críticas por sobreetiquetar imágenes mínimamente editadas, lo que derivó en una actualización que redujo la visibilidad de la etiqueta en muchos casos (Meta, 2024; Malik, 2024). Concretamente, en el caso de los reels de Instagram y tal como se implementó en este estudio, el ícono consiste en una etiqueta textual de pequeño tamaño superpuesta en la esquina inferior izquierda de la imagen, que explicita "Información de IA" en tipografía reducida de color blanco sobre fondo semitransparente, sin ningún elemento visual adicional que lo destaque, es decir, es un ícono exclusivamente textual. Su tamaño y ubicación están diseñados para cumplir los estándares mínimos de divulgación de Meta, lo que explica que tienda a pasar inadvertido para una proporción relevante de usuarios. Es posible que, a consecuencia de ello, el ícono haya adquirido en la experiencia de los usuarios un carácter más bien técnico o burocrático, sin una función comunicativa activadora. El resultado, o más

bien, el “no resultado” obtenido se encuentra en línea con el Modelo Heurístico-Sistemático (HSM; Chaiken, 1987). En contextos de consumo rápido de contenidos —como el desplazamiento continuo en Instagram o TikTok— los usuarios procesan la información mediante “atajos” cognitivos, más que a través de análisis profundos del mensaje. En esas condiciones, una señal de baja prominencia como el ícono estudiado, difícilmente alcanza el umbral de relevancia perceptual necesario para interrumpir el procesamiento heurístico habitual y desencadenar una evaluación sistemática del emisor. Este mecanismo explica por qué señales simples como el número de seguidores pueden afectar la percepción de popularidad del influencer (De Veirman, Cauberghe & Hudders, 2017), mientras que un ícono pequeño ubicado en un margen de la imagen puede pasar inadvertido, sin dejar huella evaluativa.

Al contrario, el resultado se encuentra en contradicción con la Teoría de la Violación de las Expectativas (EVT; Burgoon, 1978), la cual predice que cuando una señal produce una violación de expectativas suficientemente relevante, activa procesos de reevaluación del emisor. Que esto no haya ocurrido en este estudio podría explicarse por dos vías. Primero, es posible que la violación de expectativas producida por el ícono haya sido de baja intensidad: los participantes jóvenes, altamente expuestos a redes sociales y con uso frecuente de herramientas de IA (Activa Research, 2024), pueden haber integrado en sus esquemas la posibilidad de que el contenido de influencers sea editado o generado artificialmente, haciendo que el ícono no resulte especialmente sorprendente. Segundo, la EVT distingue entre violaciones positivas y negativas, dependiendo de la valencia del comunicador: si el emisor es percibido positivamente, incluso una violación puede ser interpretada de manera favorable (Burgoon, 1978). En este caso, el ícono podría haber sido leído como una señal de honestidad, antes que una señal de artificialidad, neutralizando así su potencial efecto negativo. Esta interpretación se articula, además, con lo planteado por el Persuasion Knowledge Model (PKM; Friestad & Wright, 1994), en tanto la respuesta del receptor no depende únicamente de la presencia de una señal, sino de la forma en que esta es decodificada dentro del contexto comunicacional.

Desde el Persuasion Knowledge Model (PKM; Friestad & Wright, 1994), para que el ícono active escepticismo, los usuarios deben reconocerlo como una señal de intento persuasivo. Si

el ícono es interpretado como un dato técnico o una convención institucional —similar a los sellos de calidad en etiquetas de productos alimenticios—, el mecanismo de activación del conocimiento persuasivo no se dispara. Kirmani y Campbell (2008) advierten que este proceso depende de la interpretación del receptor, no de la intención del emisor: si el usuario no percibe que el ícono marca un intento de influencia, la resistencia cognitiva no se activa.

A ello se suma que la advertencia utilizada en este estudio se presentaba en una ubicación de baja prominencia visual y con una formulación breve, integrada de manera discreta al entorno de la publicación, lo que reduce significativamente su capacidad de captar atención. En un contexto de navegación rápida y consumo fragmentado de contenido, como ocurre en plataformas como Instagram o TikTok, este tipo de señales tiende a pasar inadvertido y no alcanza a generar la suspicacia necesaria para activar una evaluación más crítica del emisor. Más que funcionar como una alerta, el ícono puede haber sido percibido como un elemento accesorio del diseño de la plataforma, sin relevancia suficiente para modificar la percepción de credibilidad del influencer.

En el contexto de consumo cotidiano de redes sociales, donde los usuarios están habituados a convivir con contenido patrocinado, filtrado y editado (Arriagada & Valderrama, 2023), es esperable que el ícono de IA sea procesado como un dato más del entorno visual, sin activar defensas evaluativas.

Desde un punto de vista ético o regulatorio, este hallazgo no debería interpretarse como evidencia de que la transparencia sobre contenido generado por IA sea irrelevante. Más bien, lo que el resultado sugiere más precisamente, es que la forma concreta en que esa transparencia se comunica visualmente es insuficiente para producir los efectos que los marcos regulatorios le atribuyen. Esta distinción es crucial para el debate actual: el AI Act de la Unión Europea, que comenzará a aplicarse plenamente en agosto de 2026, exige que los usuarios sean informados de que el contenido fue generado por IA, pero no prescribe un formato único de identificación visual (European Commission, 2024a). Los resultados de este estudio aportan evidencia empírica sobre los no efectos que arroja un formato específico de etiquetado, lo cual puede ser un insumo valioso para definir estándares más efectivos.

H2: Las imágenes generadas por IA no obtuvieron menores puntajes que las de influencers humanos

La segunda hipótesis esperaba que los influencers generados por IA fueran evaluados con menores niveles de credibilidad, autenticidad y confianza en comparación con los influencers humanos. Esta hipótesis tampoco fue confirmada. Más aún, los datos descriptivos muestran una tendencia que va en la dirección contraria a lo esperado: en los grupos sin ícono, el influencer de IA (G4) obtuvo puntajes de autenticidad y confianza ligeramente superiores a los del influencer humano sin ícono.

Este resultado se aleja de lo encontrado por Morosoli et al. (2024) y Schilke et al. (2025), quienes documentaron que el etiquetado de contenido como generado por IA reduce la credibilidad y la confianza percibidas. Sin embargo, cabe destacar que estas investigaciones trabajaron con condiciones en las que la artificialidad era explícita o más evidente. En cambio, los estímulos del presente estudio fueron diseñados para maximizar la similitud visual entre la imagen humana y la de IA, replicando el mismo encuadre, composición y estilo, en un afán de no agregar factores extras que pudiesen confundir a los usuarios. En ausencia del ícono, los participantes probablemente no detectaron diferencias entre ambas condiciones. Esto es coherente con lo que señalan Lee et al. (2024): cuando el contenido de IA tiene altos grados de realismo visual, las audiencias tienden a evaluarlo de manera similar al contenido humano, porque los atributos que usan para juzgar al emisor —apariencia, coherencia comunicacional, atractivo percibido— están igualmente presentes en ambas imágenes. En este contexto, la diferencia no radica necesariamente en si la imagen fue generada por una persona o por inteligencia artificial, sino en la capacidad de la audiencia para percibir esa distinción. Cuando dicha diferencia no es evidente, la evaluación tiende a desplazarse desde el origen del contenido hacia las características visibles del emisor.

Desde la perspectiva de la credibilidad de la fuente (Ohanian, 1990), lo que los participantes evalúan no es el origen real de la imagen, sino los atributos percibidos del emisor: su experiencia aparente, su confiabilidad y su atractivo. Si esos atributos están presentes en ambas condiciones de forma equivalente —lo que el diseño de estímulos buscó garantizar—, no hay razón teórica para esperar diferencias en la evaluación. Este resultado también es coherente con lo planteado por Willemsen et al. (2024), quienes muestran que los influencers

de IA con alto grado de antropomorfismo pueden alcanzar niveles similares de credibilidad que los humanos, especialmente cuando los indicadores de artificialidad no son evidentes. En ese sentido, los resultados de este estudio van en línea con esa evidencia: el hiperrealismo alcanzado por los sistemas de generación de imágenes actuales hace que la distinción perceptual humano versus IA se vuelva prácticamente inoperante para las audiencias.

Una lectura distinta ofrece la teoría del Valle Inquietante (Mori, 1970; Gutiérrez Aguilar, 2021), que predice que la similitud casi humana —pero imperfecta— de un agente artificial genera respuestas afectivas negativas como extrañeza o rechazo. Sin embargo, ese efecto presupone que el usuario detecta la imperfección; si la imagen artificial es visualmente indistinguible de una real, el mecanismo no se activa. Es posible que los estímulos utilizados en este estudio hayan superado el "valle" y alcanzado un nivel de realismo en el que la respuesta afectiva es equivalente a la producida por una imagen humana. Esta hipótesis es consistente con los desarrollos más recientes de la teoría, que enfatizan que el efecto de extrañeza depende de la percepción de ambigüedad categorial entre lo humano y lo no humano (Gutiérrez Aguilar, 2021): cuando esa ambigüedad no existe porque la imagen se percibe simplemente como humana, el efecto desaparece.

Adicionalmente, la literatura sobre la heurística de máquina (Sundar, 2008; Sundar & Kim, 2019) plantea que algunos usuarios atribuyen a los sistemas automatizados mayor objetividad o precisión que a los emisores humanos, lo que en ciertas condiciones puede elevar la evaluación del agente artificial. Aunque esta heurística no fue directamente medida en el estudio, podría contribuir a explicar por qué la condición IA sin ícono obtuvo puntajes descriptivamente similares o ligeramente superiores a la condición humana sin ícono: los participantes podrían haber procesado la imagen de IA con mayor neutralidad evaluativa, sin proyectar sobre ella las mismas sospechas de intención persuasiva que se activan frente a los influencers humanos (Friestad & Wright, 1994).

Este resultado tiene implicancias relevantes tanto para la teoría como para la práctica. A nivel teórico, cuestiona la aplicabilidad directa del Valle Inquietante en contextos de imágenes hiperrealistas, y sugiere que los marcos sobre credibilidad de la fuente (Ohanian, 1990) deben incorporar el nivel de detectabilidad de la artificialidad como variable moderadora clave. A nivel práctico, el hallazgo plantea una pregunta urgente para el debate sobre desinformación

y manipulación digital: si el contenido de IA es visualmente indistinguible del real, y si además el ícono de transparencia no produce efectos perceptuales significativos, las audiencias carecen de herramientas efectivas para evaluar el origen del contenido que consumen. Esta situación es especialmente relevante en Chile, donde la penetración de redes sociales alcanza aproximadamente el 81,7% de la población (DataReportal, 2026) y donde no existe aún una regulación específica sobre etiquetado de contenido generado por IA.

H3: El patrón de interacción fue contrario al esperado

La tercera hipótesis predecía que el efecto negativo del ícono de transparencia sería mayor cuando la imagen correspondiera a un influencer humano que cuando correspondiera a uno generado por IA. El razonamiento detrás de esta predicción descansaba en la EVT (Burgoon, 1978): si el usuario espera una representación auténtica de una persona real y el ícono le indica que la imagen es artificial, la violación de expectativas debería ser más intensa que en el caso de un emisor ya identificado como de IA.

La hipótesis no fue confirmada. No obstante, el patrón descriptivo encontrado en la variable confianza resulta particularmente informativo: el ícono produjo una diferencia de apenas 0.08 puntos en la condición humana (G1 vs. G2: M = 3.23 vs. 3.31), mientras que en la condición de IA la diferencia fue de 0.72 puntos (G3 vs. G4: M = 2.98 vs. 3.70). Es decir, el ícono parece haber afectado más la confianza cuando la imagen ya era de un influencer de IA —exactamente al revés de lo que H3 predecía.

Una interpretación posible es que el ícono actúa como señal de confirmación o de disonancia según el contexto. Cuando la imagen corresponde a un influencer humano y aparece el ícono de IA, puede generarse una ambigüedad cognitiva: "¿la imagen es de un humano o de IA?". Esta confusión podría haber moderado el impacto del ícono sobre la confianza, diluyendo su efecto negativo. En cambio, cuando la imagen ya era de IA y el ícono lo confirma, la señal es coherente con la categoría previamente activada, lo que podría intensificar la penalización sobre la confianza al reforzar la percepción de artificialidad. Esta lógica es consistente con el HSM (Chaiken, 1987): cuando una señal es ambigua respecto al tipo de emisor, el procesamiento heurístico no se orienta de manera clara, mientras que cuando la señal confirma una categoría ya activada, su impacto sobre el juicio es más directo.

Desde el PKM (Friestad & Wright, 1994), también es posible que en la condición humana + ícono el receptor haya interpretado la etiqueta como un error técnico o un marcador de edición menor —dado que Meta también aplica el ícono a imágenes humanas editadas con herramientas de IA (Meta, 2024)— reduciendo así su activación de conocimiento persuasivo. En cambio, en la condición IA + ícono, la etiqueta confirmaría que el emisor es completamente artificial, activando más fuertemente las defensas evaluativas del receptor (Kirmani & Campbell, 2008).

Este resultado invertido respecto a H3 ofrece una lección metodológica importante: las hipótesis de interacción en diseños experimentales con contenido generado por IA podrían considerar que la forma en que los usuarios categorizan mentalmente al emisor —antes de recibir cualquier señal explícita— es una variable mediadora no trivial. La literatura sobre credibilidad de la fuente (Ohanian, 1990) sugiere que los usuarios construyen sus evaluaciones a partir de una combinación de señales del emisor y del mensaje; lo que este estudio agrega es que la coherencia o incoherencia entre esas señales puede moderar el efecto de la transparencia de maneras no lineales. Este resultado sugiere la conveniencia de explorar, en estudios futuros, cómo los participantes categorizan espontáneamente al emisor como humano o artificial antes de la exposición a señales explícitas de transparencia.

7.1. Síntesis teórica integrada: el rol diferenciado de las cinco perspectivas

Los resultados obtenidos permiten ordenar, con mayor precisión, el aporte específico de cada una de las cinco perspectivas teóricas que sustentaron este estudio, más allá de su presentación individual en el marco teórico. Las tres teorías centrales —EVT, HSM y Credibilidad de la Fuente— explican de manera directa el patrón de no efectos observado: la señal de transparencia no alcanzó la prominencia visual necesaria para activar una violación de expectativas relevante (EVT), fue probablemente procesada bajo la vía heurística propia del consumo rápido de redes sociales (HSM), y los atributos de fuente —experiencia, honestidad, atractivo— permanecieron equivalentes entre condiciones porque el estímulo de IA fue diseñado para maximizar el realismo (Ohanian, 1990). Estas tres teorías, en conjunto, son suficientes para explicar por qué H1 y H2 no se confirmaron.

Las dos teorías de apoyo, en cambio, cumplen un rol distinto y más acotado: no explican la ausencia de efecto principal, sino que ofrecen una lectura complementaria del único patrón descriptivo relevante que sí emergió —la caída de 0.72 puntos en confianza dentro de la condición de IA con ícono (H3)—. El PKM (Friestad & Wright, 1994) permite entender por qué ese efecto se concentró específicamente en la variable confianza y no en credibilidad ni autenticidad: la confianza es la variable más directamente vinculada a la disposición a aceptar un mensaje sin cuestionarlo, y por tanto la más sensible a la activación del conocimiento persuasivo cuando una señal confirma —en lugar de contradecir— la categoría de emisor ya percibida. La teoría del Valle Inquietante, por su parte, no logra explicar un rechazo generalizado hacia el influencer de IA —que de hecho no fue penalizado en ausencia del ícono—, pero sí ofrece un marco plausible para el efecto de interacción encontrado: cuando la etiqueta confirma explícitamente la artificialidad de una imagen ya percibida como casi-humana, podría reactivarse la ambigüedad categorial que la teoría identifica como gatillante de la respuesta afectiva negativa (Gutiérrez Aguilar, 2021).

En consecuencia, la contribución de este estudio no reside en haber puesto a prueba cinco teorías de manera equivalente, sino en haber podido distinguir empíricamente su alcance explicativo: tres de ellas dan cuenta del fenómeno principal (la inefectividad del ícono), y dos de ellas —operando como mecanismos de segundo orden— dan cuenta del único matiz relevante detectado en los datos. Esta jerarquización explícita responde a la necesidad, señalada por la literatura metodológica en diseños multiteóricos, de especificar el nivel de análisis en el que cada marco opera, evitando que la multiplicidad teórica derive en yuxtaposición sin integración (Shadish et al., 2002).

Sobre el significado de los efectos nulos y los tamaños de efecto encontrados

Una cuestión metodológica central en la interpretación de este estudio es el significado de los resultados no significativos. Cohen (1992) advierte que un resultado no significativo no es sinónimo de ausencia de efecto: puede indicar que el efecto existe pero su magnitud es demasiado pequeña para ser detectada con el tamaño muestral disponible, o que las condiciones experimentales no generaron suficiente varianza para producir diferencias detectables. En este estudio, los valores de Pillai's Trace fueron mínimos en los tres efectos

evaluados, lo que sugiere que, de existir diferencias reales entre los grupos, su magnitud sería muy pequeña (Cohen, 1992; Field, 2018).

Esta información es en sí misma relevante: indica que, bajo condiciones similares a las de este estudio —imágenes hiperrealistas, ícono de baja visibilidad, muestra de jóvenes con alta exposición a redes sociales—, los efectos esperados sobre credibilidad, autenticidad y confianza serían difícilmente detectables a nivel individual. Este hallazgo contribuye a calibrar expectativas sobre el impacto real del ícono en la práctica cotidiana de consumo de redes sociales y dialoga con la evidencia de Koning y Voorveld (2025) sobre la sensibilidad contextual de los efectos de las divulgaciones de IA.

Adicionalmente, la asimetría en el tamaño de los grupos experimentales —producto del criterio de exclusión aplicado a los participantes que no detectaron el ícono, siguiendo el procedimiento estándar descrito por Oppenheimer et al. (2009)— limitó el poder estadístico del análisis. Los grupos con ícono (G1 y G3) quedaron con solo 30 participantes cada uno, frente a los 55 y 86 participantes de los grupos sin ícono. Esta disparidad aumenta la varianza de los estimadores en el MANOVA (Tabachnick & Fidell, 2013; Hair et al., 2019) y reduce la probabilidad de detectar diferencias reales, especialmente en análisis multivariados. El resultado podría verse como un límite del poder estadístico antes que como evidencia definitiva de ausencia de efecto. En ese sentido, el análisis exploratorio de la variable confianza —que se acercó al umbral de significación sugiere que esta variable podría ser la más sensible a las condiciones experimentales y merece ser estudiada con mayor poder estadístico en investigaciones futuras. Si bien una alternativa metodológica habría sido igualar artificialmente los tamaños muestrales mediante submuestreo aleatorio, esta opción implicaría la pérdida de información y una reducción adicional del poder estadístico. Por esta razón, se optó por mantener todos los casos disponibles, priorizando el uso de estadísticos multivariados robustos al desbalance.

8. Limitaciones, aprendizajes y sugerencias para futuras investigaciones

Este estudio no está exento de limitaciones, las cuales dejan aprendizajes y consejos concretos para quienes investiguen el impacto de la transparencia en IA sobre percepciones

comunicacionales, especialmente en la temática de la persuasión de influencers en redes sociales en el contexto latinoamericano.

En primer lugar, la tasa de exclusión por falta de detección del ícono fue alta: de 319 participantes originales, 118 debieron ser excluidos porque no reportaron haber visto el ícono en la imagen, a pesar de haber sido asignados a una condición con ícono. Esto implica que más de la mitad de los participantes en esas condiciones no detectaron el estímulo central del experimento. Si bien el criterio de exclusión aplicado sigue el procedimiento estándar para la verificación de manipulación en estudios experimentales (Oppenheimer et al., 2009; Rodríguez Hidalgo, 2018), la magnitud de la exclusión es en sí misma un hallazgo: sugiere que el ícono tiene una visibilidad real mucho menor a la esperada, incluso en condiciones controladas de observación. Investigaciones futuras deberían considerar manipulaciones de mayor prominencia visual —por ejemplo, etiquetas con texto explícito, mayor tamaño o contraste— y probar distintos formatos antes de implementar el diseño experimental.

Asimismo, el estudio mostró un marcado desbalance entre los grupos resultante del criterio de verificación de manipulación: los grupos con ícono quedaron con solo 30 participantes cada uno, mientras que los grupos sin ícono contaron con 86 y 55 participantes respectivamente. Si bien, este tipo de desbalance es relativamente tolerado por análisis como MANOVA, el principal desafío radica en el tamaño muestral reducido de las condiciones con ícono, que podría ser insuficiente para detectar efectos de magnitud pequeña, contribuyendo así a la falta de significación estadística observada.

En una línea distinta, pero conceptualmente relacionada, investigaciones previas han señalado que determinados resultados no esperados pueden deberse a características del estímulo o del entorno experimental, más que a la inexistencia del efecto. En particular, los problemas de validez ecológica y de realismo del estímulo han sido ampliamente documentados como factores que pueden modular las respuestas de los participantes en contextos experimentales (Shadish, Cook, & Campbell, 2002).

En segundo lugar, dado que en investigaciones perceptuales suelen esperarse efectos de magnitud pequeña (Cohen, 1992), resulta recomendable realizar un análisis de potencia a priori con el fin de asegurar una sensibilidad estadística adecuada. En el presente estudio, el diseño cuasi experimental con asignación no aleatoria —justificado por el contexto de

aplicación en línea— dio lugar a tamaños de grupo desiguales, lo que, junto con el tamaño muestral disponible, pudo haber reducido el poder estadístico del análisis MANOVA, particularmente en la detección de efectos pequeños (Tabachnick & Fidell, 2013). Futuros estudios deberían apuntar a grupos de al menos 60-80 participantes por celda para tener potencia suficiente en diseños 2×2 multivariados.

En tercer lugar, el alto nivel de realismo de los estímulos visuales podría haber dificultado que los participantes distinguieran entre la imagen humana y la de IA en ausencia del ícono, lo que posiblemente redujo la diferenciación perceptual entre las condiciones experimentales. Estudios previos que sí encontraron diferencias entre influencers humanos y de IA trabajaron con agentes claramente no humanos o con menor grado de antropomorfismo (Appel, Müller & Stiglbauer, 2020; Lee et al., 2024).

En cuarto lugar, las escalas utilizadas en este estudio mostraron buena consistencia interna (α = .817 para credibilidad, .866 para autenticidad, .866 para confianza), lo que confirma su adecuación para medir estos constructos en muestras chilenas de jóvenes adultos, en línea con los estándares recomendados por Nunnally y Bernstein (1994). Sin embargo, dado que la confianza fue la variable que mostró las tendencias descriptivas más interesantes, sería valioso explorar en futuros estudios medidas de reacción más inmediata —como escalas de valencia afectiva o *arousal*— que capturen la dimensión emocional de la respuesta al contenido artificial, dimensión que las escalas Likert de acuerdo/desacuerdo no recogen plenamente. Asimismo, futuras investigaciones podrían considerar la medición de la confianza como un constructo independiente, fuera de un enfoque multivariado, e incorporar un mayor número y diversidad de ítems —incluyendo indicadores implícitos— con el fin de captar con mayor sensibilidad sus variaciones.

En quinto lugar, investigaciones futuras podrían beneficiarse de incorporar variables moderadoras como la alfabetización digital, las actitudes previas hacia la inteligencia artificial y el nivel de conocimiento sobre la existencia de influencers generados por IA. El estudio de Activa Research (2024), señala que en Chile existe un 38% de la población con poco o ningún conocimiento sobre cómo la IA interactúa en su vida cotidiana, y que la confianza en la IA es moderada: solo un 29,2% declara confiar en ella. Es esperable que usuarios con mayor conocimiento respondan de manera diferente al ícono que usuarios menos informados, y que

este efecto moderador sea relevante tanto académica como normativamente. Li & See-To (2024) destacan precisamente que el efecto de la transparencia sobre la confianza varía según el nivel de conocimiento previo del usuario sobre sistemas de IA, lo que justificaría incluir esta variable en estudios futuros sobre el tema.

En sexto lugar, más allá de las recomendaciones metodológicas para investigaciones futuras, los hallazgos de este estudio permiten formular sugerencias concretas de diseño para plataformas digitales y organismos reguladores. En primer lugar, la evidencia de que 118 de 319 participantes (37%) no detectaron el ícono incluso cuando se les instruyó observar la imagen con atención durante 15 segundos, sugiere que un formato puramente textual, de bajo contraste y ubicado en una esquina secundaria de la publicación, no cumple un estándar mínimo de visibilidad. Un diseño más efectivo debería considerar al menos uno de los siguientes elementos: (a) mayor contraste cromático respecto del fondo de la imagen, en lugar de texto blanco semitransparente; (b) un tamaño mínimo relativo al área total del estímulo, en vez de una etiqueta de pocos píxeles; o (c) un formato de interrupción breve —por ejemplo, un aviso de un segundo antes de mostrar el contenido— en lugar de una superposición pasiva que compite con otros elementos visuales de la publicación (corazones, contador de comentarios, nombre de usuario).

Por otro lado, el hallazgo de que el ícono sí produjo una caída descriptiva relevante en confianza cuando confirmaba una categoría de emisor ya percibida como artificial (G3), pero no cuando introducía ambigüedad sobre un emisor percibido como humano (G1), sugiere que el momento de la divulgación importa tanto como su formato. Una etiqueta aplicada de manera posterior a la exposición inicial del contenido —como ocurre actualmente en Instagram, donde el ícono aparece integrado en la publicación ya visible— permite que el usuario procese primero la imagen bajo su categoría heurística inicial (humano vs. artificial) antes de recibir la señal de transparencia, reduciendo su capacidad de reconfigurar esa evaluación. Una divulgación anticipada —por ejemplo, previa a la reproducción del contenido, como ya aplican TikTok o YouTube en ciertos formatos— podría producir un efecto más consistente con el principio de transparencia informada que persigue la regulación.

Estas sugerencias no deben leerse como recomendaciones definitivas, dado que no fueron puestas a prueba experimentalmente en este estudio, sino como hipótesis de diseño derivadas directamente de los patrones observados, que podrían orientar tanto el trabajo de las plataformas como futuras rondas de investigación aplicada en este campo.

9. Conclusiones

Esta investigación tuvo como objetivo analizar si la presencia de un ícono de transparencia que indica el origen generado por inteligencia artificial y el tipo de imagen utilizado — influencer humano vs. generado por IA— producían diferencias en la percepción de credibilidad, autenticidad y confianza de usuarios jóvenes chilenos. A partir de un diseño cuasi experimental factorial 2×2 con 201 participantes y un análisis MANOVA, el estudio no encontró evidencia estadística significativa que indique que el ícono afecte la percepción de estas variables en una muestra de jóvenes chilenos. Sin embargo, siguiendo la recomendación de Cohen (1992) de extraer valor informativo de los efectos nulos cuando estos se producen bajo condiciones metodológicamente controladas, los hallazgos de este estudio permiten formular conclusiones relevantes tanto para la investigación en comunicación digital como para el debate sobre regulación y diseño de transparencia en IA.

La primera conclusión más central de este estudio es que el ícono de transparencia de IA, en su forma actual de implementación en plataformas como Instagram, no altera de manera perceptible la evaluación que los usuarios jóvenes hacen del contenido ni del emisor en términos de credibilidad, autenticidad o confianza. Esta conclusión tiene una doble lectura. Por un lado, podría interpretarse como una buena noticia para los emisores que utilizan IA en sus contenidos: la divulgación, al menos en este formato, no penaliza la percepción de los usuarios sobre el influencer. Por otro lado, y desde una perspectiva ética y regulatoria, el hallazgo es preocupante: un mecanismo de transparencia que no produce efectos sobre la audiencia es, en la práctica, un mecanismo que no cumple su función. Li & See-To (2024) subrayan que la transparencia sobre el uso de IA solo afecta la confianza del usuario cuando es procesada de manera activa; si la señal pasa inadvertida o es cognitivamente neutralizada, no produce los efectos esperados. Los datos de este estudio sugieren que eso es precisamente lo que ocurre con el ícono actual. Los resultados de esta investigación indican que el ícono de

transparencia, en su configuración actual, no logra activar un procesamiento consciente por parte de la audiencia ni generar un efecto significativo en la evaluación del contenido o del emisor.

En este sentido, el aporte más relevante de esta investigación no es confirmar o refutar hipótesis sobre los efectos del ícono, sino desplazar la pregunta de investigación hacia un nivel más profundo: no si etiquetar el contenido generado por IA, sino cómo hacerlo de manera que produzca efectos reales sobre la audiencia. Esta distinción es relevante para los marcos regulatorios en desarrollo. El AI Act de la Unión Europea (European Commission, 2024a), que comenzará a aplicarse plenamente en agosto de 2026, exige transparencia sobre el uso de IA en contenidos que puedan inducir a error, pero no prescribe un formato único de etiquetado. Los resultados de este estudio aportan evidencia empírica sobre los límites de un formato específico y pueden servir de insumo para el diseño de estándares más efectivos, tanto a nivel internacional como en el debate legislativo chileno. En este ámbito, aunque actualmente no existe una regulación vigente que obligue al etiquetado de contenidos generados mediante IA, el país cuenta con avances en tramitación legislativa, como la iniciativa ingresada en 2023 y su posterior integración al Boletín N.º 16821-19, actualmente en segundo trámite en el Senado (Aedo et al., 2023). Este escenario refuerza la pertinencia de generar evidencia aplicada que contribuya a orientar decisiones regulatorias en curso.

En particular, si el proyecto refundido en el Boletín N° 16821-19 avanza hacia una obligación de etiquetado similar a la contemplada en el artículo 50 del Reglamento europeo de IA —que exige informar al usuario sin imponer un formato visual único—, los resultados de este estudio son un antecedente empírico directo para esa discusión: una obligación de transparencia que se limite a exigir la mera presencia de una etiqueta, sin especificar criterios mínimos de tamaño, contraste o momento de exposición, corre el riesgo de traducirse en un cumplimiento formal sin efecto perceptual real sobre la audiencia, tal como ocurrió con el ícono de Meta evaluado en este estudio. Este riesgo es especialmente relevante en el contexto legislativo chileno, donde la discusión se encuentra aún en una etapa de definición de estándares técnicos y donde la evidencia comparada sobre la efectividad de distintos formatos de divulgación es todavía escasa. En este sentido, este estudio aporta un antecedente concreto para que la futura normativa —o su reglamento de aplicación— no se

limite a exigir la existencia de una señal de transparencia, sino que establezca parámetros mínimos de visibilidad y procesamiento cognitivo que aseguren su efectividad comunicacional, evitando que la transparencia se convierta en un ejercicio de cumplimiento nominal antes que en una herramienta real de protección al usuario.

Una segunda conclusión relevante es que, bajo condiciones de alta fidelidad visual, las imágenes de influencers generadas por IA no son evaluadas de manera significativamente diferente a las de influencers humanos. Los sistemas de generación de imágenes contemporáneos han alcanzado un nivel de realismo tal que las audiencias, incluso jóvenes con alta exposición a redes sociales, no perciben diferencias en ausencia de señales explícitas (Willemsen et al., 2024; Lee et al., 2024). Este hallazgo tiene implicancias directas para los debates sobre autenticidad y desinformación en entornos digitales: si el contenido artificial es visualmente indistinguible del real —y si además el ícono de transparencia no produce efectos significativos sobre las percepciones— las audiencias carecen, en la práctica, de herramientas efectivas para evaluar el origen del contenido que consumen. Esta situación adquiere especial relevancia en Chile, donde la penetración de redes sociales alcanza aproximadamente el 81,7% de la población (DataReportal, 2026) y donde los jóvenes de 18 a 29 años utilizan las plataformas digitales como principal fuente de información (Feedback Research & Universidad Diego Portales, 2024).

Una tercera conclusión de interés es el patrón descriptivo observado en la variable confianza dentro de los grupos de influencer generados por IA: el ícono estuvo asociado a una caída de 0.72 puntos respecto a la condición sin ícono, mientras que en los grupos de influencer humano la diferencia fue de apenas 0.08 puntos. Si bien este efecto de interacción no alcanzó significación estadística (p = .235), es el patrón más coherente con las predicciones teóricas y constituye la señal descriptiva más informativa del estudio. Desde la Teoría del Conocimiento de la Persuasión (Friestad & Wright, 1994) y el Modelo Heurístico-Sistemático (Chaiken, 1987), es plausible que el ícono produzca un efecto más intenso cuando confirma una categoría ya activada (agente artificial) que cuando introduce ambigüedad sobre un emisor percibido como humano. Investigaciones futuras con mayor poder estadístico y diseños que manipulen explícitamente la presentación del origen del emisor —asegurando su identificación como humano o artificial— podrán determinar si este patrón es sistemático, y

si la confianza opera como variable bisagra entre la percepción de artificialidad y la evaluación global del mensaje (Gohil et al., 2025).

Desde el punto de vista teórico, este estudio contribuye a matizar la aplicabilidad de tres marcos conceptuales en el contexto específico de los influencers de IA hiperrealistas. Primero, la teoría del Valle Inquietante (Mori, 1970; Gutiérrez Aguilar, 2021) predice respuestas afectivas negativas ante agentes casi humanos; sin embargo, cuando la imagen artificial supera el umbral de detectabilidad visual, el mecanismo de extrañeza no se activa, lo que limita la aplicabilidad directa de la teoría en este contexto. Segundo, la EVT (Burgoon, 1978) predice que las violaciones de expectativas activan reevaluaciones del emisor, pero ese proceso requiere que la señal sea suficientemente saliente: cuando el ícono es percibido como una convención técnica de baja relevancia, la violación de expectativas es demasiado leve para producir el efecto esperado. Tercero, el HSM (Chaiken, 1987) explica por qué en entornos de consumo rápido las señales de baja prominencia son procesadas heurísticamente, sin activar juicios evaluativos profundos sobre el emisor. En conjunto, estos marcos permiten comprender el resultado nulo no como un fracaso teórico, sino como una evidencia de los límites de los estímulos comunicacionales de baja intensidad en contextos de alta saturación mediática.

A nivel metodológico, el estudio valida el uso de escalas de credibilidad, autenticidad y confianza en muestras chilenas de jóvenes adultos, con resultados de consistencia interna satisfactorios ($\alpha \geq .817$), en línea con los criterios de Nunnally & Bernstein (1994). También documenta los desafíos propios de los diseños experimentales en línea, con estímulos visuales de baja visibilidad, en particular la alta tasa de exclusión por falta de detección del ícono —118 participantes sobre 319 originales—, que afectó el poder estadístico del análisis y que es en sí misma un hallazgo sobre la visibilidad real del etiquetado de IA en condiciones naturales de consumo.

En síntesis, este estudio concluye que el principio ético de la transparencia en IA es válido y necesario, pero que su traducción comunicacional actual —representada por un ícono visual pequeño y de baja prominencia— es insuficiente para producir los efectos perceptuales que se le atribuyen normativamente. Esta conclusión es, en sí misma, un aporte al debate: no pone en duda el valor de la transparencia, sino que evidencia la brecha entre la transparencia

como principio y la transparencia como práctica comunicacional efectiva. Cerrar esa brecha —diseñando formatos de etiquetado que sí produzcan efectos reales sobre las percepciones de los usuarios— es uno de los desafíos más urgentes para quienes investigan, regulan y practican la comunicación estratégica digital en la era de la inteligencia artificial generativa.

Finalmente, este estudio abre una agenda de investigación relevante para el contexto latinoamericano. Queda por responder qué formatos de etiquetado para advertencias de uso de IA son efectivos para usuarios con distintos niveles de alfabetización digital (Activa Research, 2024; Li & See-To, 2024). Junto a ello, sería relevante investigar si es que existen diferencias generacionales o socioeconómicas en la respuesta a las señales de transparencia IA, y cómo interactúa la variable de confianza institucional en las plataformas (Arriagada & Valderrama, 2023) con la confianza en los influencers individuales ante el contenido generado por IA. Abordar estas preguntas con diseños experimentales de mayor poder estadístico, muestras más amplias y diversas, avanzaría el conocimiento de manera relevante ante una problemática urgente. Además, se podría considerar una colaboración entre academia, plataformas y organismos reguladores, para avanzar hacia políticas de transparencia que sean no solo éticamente correctas, sino también comunicacionalmente efectivas.

Referencias

Abidin, C. (2016). “Aren’t these just young, rich women doing vain things online?” Influencer selfies as subversive frivolity. *Social media + Society*, *2*(2). https://doi.org/10.1177/2056305116641342

Activa Research. (2024). *Estudio “Los chilenos y la inteligencia artificial”*. https://chile.activasite.com/estudios/estudio-los-chilenos-y-la-inteligencia-artificial/

Aedo, E., Lagomarsino, T., Medina, K., Mellado, C., Olivera, E., Oyarzo, R., Santibáñez, M., & Venegas, N. (2023). *Proyecto de ley que regula los sistemas de inteligencia artificial, la robótica y las tecnologías conexas, en sus distintos ámbitos de aplicación (Boletín N° 15869-19)*. Cámara de Diputadas y Diputados de Chile. https://www.camara.cl/legislacion/proyectosdeley/tramitacion.aspx?prmID=16416&prmBOLETIN=15869-19

Appel, M., Müller, J., & Stiglbauer, B. (2020). Human and computer-generated faces: The uncanny valley revisited. *Computers in Human Behavior*, *111*, 106416. https://doi.org/10.1016/j.chb.2020.106416

Arnett, J. J. (2000). Emerging adulthood: A theory of development from the late teens through the twenties. *American Psychologist*, *55*(5), 469–480. https://doi.org/10.1037/0003-066X.55.5.469

Arnett, J. J., Žukauskienė, R., & Sugimura, K. (2014). The new life stage of emerging adulthood at ages 18–29 years: Implications for mental health. *The Lancet Psychiatry*, *1*(7), 569–576. https://doi.org/10.1016/S2215-0366(14)00080-7

Arriagada, A., & Ibáñez, F. (2020). “You need at least one picture daily, if not, you’re dead”: Content creators and platform dependence in the social media economy. *Social Media + Society*, *6*(3). https://doi.org/10.1177/2056305120944624

Arriagada, A., & Valderrama, M. (2023). Making social media influence calculable: Influencer agencies, algorithmic systems, and platform metrics. En P. Bouquillion, C. Ithurbide, & T. Mattelart (Eds.), *Digital Platforms and the Global South: Reconfiguring Power*

Relations in the Cultural Industries (1.a ed.). Routledge. https://doi.org/10.4324/9781003391746

Audrezet, A., de Kerviler, G., & Moulard, J. G. (2020). Authenticity under threat: When social media influencers need to go beyond self-presentation. *Journal of Business Research*, *117*, 557–569. https://doi.org/10.1016/j.jbusres.2018.07.008

Balmaceda, D., Arriagada, A., & Scherman, A. (2022). Autenticidad percibida y credibilidad en influencers digitales: Evidencia desde audiencias jóvenes en Chile. *Cuadernos.info*, *52*, 1–20.

Balmaceda, T., Paoli, M. D., & Marenco, J. (2022). *Cultura de la influencia: La fuerza suave que está moldeando una nueva sociedad*. Marea Editorial.

Bautista Jara, M., & Chávez Yépez, D. (2021). Influencers digitales y su impacto en la opinión pública juvenil. *Revista Latinoamericana de Comunicación*, *149*, 87–102.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2.a ed.). Guilford Press.

Burgoon, J. K. (1978). A communication model of personal space violations: Explication and an initial test. *Human Communication Research*, *4*(2), 129–142.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, *39*(5), 752–766. https://doi.org/10.1037/0022-3514.39.5.752

Chaiken, S. (1987). The heuristic model of persuasion. En M. P. Zanna, J. M. Olson, & C. P. Herman (Eds.), *Social influence: The Ontario symposium* (Vol. 5, pp. 3–39). Lawrence Erlbaum.

Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, *107*(6), 1753–1820.

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

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Congreso Nacional de Chile. (1970). *Ley N.º 17.336 sobre propiedad intelectual*. Biblioteca del Congreso Nacional de Chile.

Consejo de Autorregulación y Ética Publicitaria (CONAR). (2022). *Código chileno de ética publicitaria*. Autor.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555

Dabiran, E., Farivar, S., Wang, F., & Grant, G. (2024). Virtually human: Anthropomorphism in virtual influencer marketing. *Journal of Retailing and Consumer Services*, *79*, 103797. https://doi.org/10.1016/j.jretconser.2024.103797

DataReportal. (2026). *Digital 2026: Chile*. Kepios. https://datareportal.com/reports/digital-2026-chile

De Veirman, M., Cauberghe, V., & Hudders, L. (2017). Marketing through Instagram influencers: The impact of number of followers and product divergence on brand attitude. *International Journal of Advertising*, *36*(5), 798–828. https://doi.org/10.1080/02650487.2017.1348035

European Commission. (2024a). *Regulation (EU) on artificial intelligence (Artificial Intelligence Act)*. https://aiact.algolia.com/article-52/

European Commission. (2024b). *Code of practice on marking and labelling AI-generated content*. https://digital-strategy.ec.europa.eu/en/news/commission-publishes-first-draft-code-practice-marking-and-labelling-ai-generated-content

Evans, N. J., Phua, J., Lim, J., & Jun, H. (2017). Disclosing Instagram influencer advertising: The effects of disclosure language on advertising recognition, attitudes, and behavioral intent. *Journal of Interactive Advertising*, *17*(2), 138–149. https://doi.org/10.1080/15252019.2017.1366885

Feedback Research & Universidad Diego Portales. (2024). *15.ª Encuesta sobre participación, jóvenes y consumo de medios*. https://www.feedbackresearch.cl

Fernández, A. (2017). *Estudio del origen de la figura del influencer y análisis de su poder de influencia en base a sus comunidades* [Trabajo de fin de grado, Universidad Pompeu Fabra]. Repositorio UPF. https://repositori.upf.edu/handle/10230/36313

Field, A. (2018). *Discovering statistics using IBM SPSS Statistics* (5.a ed.). SAGE Publications.

Fletcher, R., Cornia, A., & Nielsen, R. K. (2019). How polarized are online and offline news audiences? A comparative analysis of twelve countries. *The International Journal of Press/Politics*, *24*(2), 154–174.

Freberg, K., Graham, K., McGaughey, K., & Freberg, L. A. (2011). Who are the social media influencers? A study of public perceptions of personality. *Public Relations Review*, *37*(1), 90–92. https://doi.org/10.1016/j.pubrev.2010.11.001

Friestad, M., & Wright, P. (1994). The persuasion knowledge model: How people cope with persuasion attempts. *Journal of Consumer Research*, *21*(1), 1–31. https://doi.org/10.1086/209380

Gohil, S. A., Nagariya, B. K., Parmar, A. U., Joshi, P. D., & Joshi, D. B. (2025). The impact of influencer authenticity on purchase intentions among Gen Z consumers. *Advances in Consumer Research*, *2*(5), 890–903.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, *27*. https://doi.org/10.48550/arXiv.1406.2661

Guerrero, M. V. (2017). *Los tuitstars: Líderes de opinión en Twitter* [Tesis de maestría, Universidad del Salvador]. RACIMO. https://racimo.usal.edu.ar/5806/

Gutiérrez Aguilar, R. (2021). Más allá del Uncanny Valley: Forma de lo humano, mímesis y expresión. *Bajo Palabra*, (28), 175–198.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8.a ed.). Cengage Learning.

Horton, D., & Wohl, R. R. (1956). Mass communication and para-social interaction: Observations on intimacy at a distance. *Psychiatry*, *19*(3), 215–229. https://doi.org/10.1080/00332747.1956.11023049.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118.

Intriago, E., Hidalgo Mayorga, M. D. L. Á., Quiñónez Bedón, M. F., & Casanova, J. (2024). Nuevos líderes de opinión en las nuevas plataformas, ámbito legal y la esfera pública. *Revista Científica Arbitrada Multidisciplinaria PENTACIENCIAS*, *6*(1), 106–115. https://doi.org/10.59169/pentaciencias.v6i1.963

Jin, S. V., Muqaddam, A., & Ryu, E. (2019). Instafamous and social media influencer marketing. *Marketing Intelligence & Planning*, *37*(5), 567–579. https://doi.org/10.1108/MIP-09-2018-0375

Ju, Y., Kim, J., & Choi, Y. K. (2024). Rise of AI influencers: A conceptualization and research agenda. *Human-centric Computing and Information Sciences*, *14*(1), 1–18. https://doi.org/10.1186/s13673-024-00404-3

Kirmani, A., & Campbell, M. C. (2008). I know what you're doing and why you're doing it: The use of the persuasion knowledge model in consumer research. En C. P. Haugtvedt, P. M. Herr, & F. R. Kardes (Eds.), *Handbook of consumer psychology* (pp. 549–573). Lawrence Erlbaum.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4.a ed.). Guilford Press.

Koning, B., & Voorveld, H. (2025). Disclaimer! This content is AI-generated: How AI disclosures influence trust in advertisements and organizations. *Journal of Interactive Advertising*. https://doi.org/10.1080/15252019.2025.2554149.

Lee, S., Kim, J., & Park, H. (2024). Human versus AI influencers: How source type and disclosure affect authenticity and persuasion. *Computers in Human Behavior*, *146*, 107768. https://doi.org/10.1016/j.chb.2023.107768.

Ley N.º 19.496. (1997). *Ley sobre protección de los derechos de los consumidores*. Biblioteca del Congreso Nacional de Chile.

Li, Y., & See-To, E. W. K. (2024). Trust in artificial intelligence–generated content: The role of transparency and user cognition. *Computers in Human Behavior*, *148*, 107889. https://doi.org/10.1016/j.chb.2023.107889.

Lou, C., & Yuan, S. (2019). Influencer marketing: How message value and credibility affect consumer trust of branded content on social media. *Journal of Interactive Advertising*, *19*(1), 58–73. https://doi.org/10.1080/15252019.2018.1533501.

Lou, C., Tan, Y., & Chen, X. (2023). Investigating consumer engagement with virtual influencers: The roles of authenticity, attractiveness, and parasocial interaction. *Journal of Retailing and Consumer Services*, *70*, 103149. https://doi.org/10.1016/j.jretconser.2022.103149.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130–149. https://doi.org/10.1037/1082-989X.1.2.130.

Malik, A. (2024, 12 de septiembre). Meta is making its AI info label less visible on content edited or modified by AI tools. *TechCrunch*. https://techcrunch.com/2024/09/12/meta-is-making-its-ai-info-label-less-visible-on-content-edited-or-modified-by-ai-tools/

McGinn, K. A. (2013). *Effective framing strategies for service advertising: The impact of narrative, rhetorical tropes and argument on consumer response across different service categories* [Tesis doctoral, City University London]. City Research Online. http://openaccess.city.ac.uk/2718/

Meta (2024a, 5 de abril). Our approach to labeling AI-generated content and manipulated media. *About Meta*. https://about.fb.com/news/2024/04/metas-approach-to-labeling-ai-generated-content-and-manipulated-media/

Meta (2024b, 6 de febrero). Labeling AI-generated images on Facebook, Instagram and Threads. *About Meta*. https://about.fb.com/news/2024/02/labeling-ai-generated-images-on-facebook-instagram-and-threads/

Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of Communication*, *60*(3), 413–439. https://doi.org/10.1111/j.1460-2466.2010.01488.x

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

Morosoli, S., van der Goot, E., Resendez, V., de Vreese, C., & Helberger, N. (2025). The transparency dilemma: An experiment on how AI disclosures affect credibility perceptions and engagement across topics. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, *8*(2), 1748–1757. https://doi.org/10.1609/aies.v8i2.36671

Moustakas, E., Lamba, N., Mahmoud, D., & Ranganathan, C. (2020). Blurring lines between fiction and reality: Perspectives of experts on marketing effectiveness of virtual influencers. *International Journal of Information Management*, *54*, 102267. https://doi.org/10.1016/j.ijinfomgt.2020.102267

Newman, N., Fletcher, R., Robertson, C. T., Eddy, K., & Nielsen, R. K. (2023). *Reuters Institute Digital News Report 2023*. Reuters Institute for the Study of Journalism, University of Oxford.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3.a ed.). McGraw-Hill.

Ohanian, R. (1990). Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. *Journal of Advertising*, *19*(3), 39–52. https://doi.org/10.1080/00913367.1990.10673191

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867–872. https://doi.org/10.1016/j.jesp.2009.03.009

Parlamento Europeo y Consejo de la Unión Europea. (2024). *Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen*

normas armonizadas en materia de inteligencia artificial (Reglamento de Inteligencia Artificial). Diario Oficial de la Unión Europea. https://www.boe.es/buscar/doc.php?id=DOUE-L-2024-81079

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. Springer.

Ricks, J. M., Jr. (2002). *The effects of strategic corporate philanthropy on consumer perceptions: An experimental assessment* [Tesis doctoral, Louisiana State University]. LSU Scholarly Repository. https://repository.lsu.edu/gradschool_dissertations/1167

Rodríguez Hidalgo, C. T. (2018). *Bits of emotion: The process and outcomes of sharing emotions online* [Tesis doctoral, Universiteit van Amsterdam]. UvA-DARE. https://dare.uva.nl

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Schilke, O., & Reimann, M. (2025). The transparency dilemma: How AI disclosure erodes trust. *Organizational Behavior and Human Decision Processes*, *188*, Artículo 104405. https://doi.org/10.1016/j.obhdp.2025.104405.

Sokolova, K., & Kefi, H. (2020). Instagram and YouTube bloggers promote it, why should I buy? How credibility and parasocial interaction influence purchase intentions. *Journal of Retailing and Consumer Services*, *53*, 101742. https://doi.org/10.1016/j.jretconser.2019.01.011.

Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. En M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 73–100). MIT Press.

Sundar, S. S., & Kim, J. (2019). Machine heuristic: When we trust computers more than humans with our personal information. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–9. https://doi.org/10.1145/3290605.3300768

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6.a ed.). Pearson.

Thomas, V. L., & Fowler, K. (2021). Close encounters of the AI kind: Use of AI influencers as brand endorsers. *Journal of Advertising*, *50*(1), 11–25. https://doi.org/10.1080/00913367.2020.1810595.

TikTok (2024). *AI-generated content policy*. https://support.tiktok.com/en/using-tiktok/creating-videos/ai-generated-content

Torronteras Manzano, R., Andaluz Antón, L., & Sacaluga Rodríguez, I. (2025). IA e influencers: El uso de la inteligencia artificial para perpetuar los estereotipos de género y la hipersexualización de la mujer. *European Public & Social Innovation Review*, *10*, 1–19. https://doi.org/10.31637/epsir-2025-1017

Tukachinsky, R. (2011). Para-social relationships: Conceptualization and measurement of parasocial relationships. *Human Communication Research*, *37*(1), 73–98. https://doi.org/10.1111/j.1468-2958.2010.01389.x

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, *6*(1). https://doi.org/10.1177/2056305120903408

Van der Meer, T. G. L. A., & Jin, Y. (2020). Seeking formula for misinformation treatment in public health crises: The effects of corrective information type and source. *Journal of Communication*, *70*(4), 550–575.

Villegas, C. (2022). Influencers y autenticidad: Una aproximación desde la comunicación estratégica digital. *Comunicación y Sociedad*, *35*(3), 45–63.

Willemsen, M. C., Neijens, P. C., & Bronner, F. (2024). When AI looks human: Effects of anthropomorphic AI-generated influencers on trust and credibility perceptions. *Journal of Interactive Marketing*, *65*, 1–15. https://doi.org/10.1016/j.intmar.2023.11.002

Zaichkowsky, J. L. (1985). Measuring the involvement construct. *Journal of Consumer Research*, *12*(3), 341–352. https://doi.org/10.1086/208520.

Anexo 1: Cuestionario completo

A continuación se presenta el cuestionario completo utilizado en el estudio, tal como fue administrado a los participantes a través de la plataforma Google Forms. El cuestionario fue aplicado en cuatro versiones paralelas correspondientes a las cuatro condiciones experimentales del diseño factorial 2×2. Cada participante respondió únicamente la versión que le fue asignada.

Consentimiento informado

Percepción y evaluación de contenidos visuales en redes sociales

Esta encuesta forma parte de una investigación sobre comunicación en redes sociales de la Universidad Adolfo Ibáñez.

Está dirigida exclusivamente a personas entre 18 y 28 años. Si usted no se encuentra dentro de este rango etario, le pedimos amablemente no continuar.

Su participación es voluntaria y anónima, y no existen respuestas correctas o incorrectas. La información recopilada será utilizada únicamente con fines académicos.

Puede retirarse en cualquier momento, sin necesidad de dar una explicación.

Responsables: Matías Vargas (matvargas@alumnos.uai.cl) y Catalina Urzúa (caturzua@alumnos.uai.cl). Profesora guía: Carmina Rodríguez (carmina.rodriguez@uai.cl).

Duración estimada: 3 a 5 minutos. Le recomendamos responder en un lugar sin distracciones.

Asignación de condición experimental

Para continuar con la encuesta, selecciona una de las siguientes opciones. No hay opciones correctas o incorrectas. Una vez que elijas una opción, responderás preguntas únicamente sobre esa opción. *

- ○ Opción A
- ○ Opción B
- ○ Opción C

○ Opción D

Presentación del estímulo

A continuación, verá una imagen similar a las que aparecen en redes sociales. Le pedimos observarla con atención durante aproximadamente 15 segundos. Una vez finalizado este tiempo, se le presentará un breve cuestionario sobre sus impresiones.

Tenga en cuenta que, al avanzar a la siguiente sección, no podrá volver a visualizar la imagen. Por este motivo, le recomendamos observarla con atención antes de continuar.

Recuerde que no existen respuestas correctas o incorrectas, solo nos interesa conocer sus impresiones.

[Imagen experimental correspondiente a la condición asignada]

Cuestionario: percepciones — Bloque 1: Credibilidad percibida

Escala de respuesta: 1 = Nada de acuerdo → 7 = Totalmente de acuerdo

1) La imagen me parece creíble. *

Escala: 1 = Nada de acuerdo → 7 = Totalmente de acuerdo

○ Nada de acuerdo

○ Muy en desacuerdo

○ En desacuerdo

○ Ni de acuerdo ni en desacuerdo

○ De acuerdo

○ Muy de acuerdo

○ Totalmente de acuerdo

2) La información que transmite la imagen me parece confiable. *

Escala: 1 = Nada de acuerdo → 7 = Totalmente de acuerdo

- ○ Nada de acuerdo
- ○ Muy en desacuerdo
- ○ En desacuerdo
- ○ Ni de acuerdo ni en desacuerdo
- ○ De acuerdo
- ○ Muy de acuerdo
- ○ Totalmente de acuerdo

3) Percibo coherencia entre la imagen y lo que comunica. *

Escala: 1 = Nada de acuerdo → 7 = Totalmente de acuerdo

- ○ Nada de acuerdo
- ○ Muy en desacuerdo
- ○ En desacuerdo
- ○ Ni de acuerdo ni en desacuerdo
- ○ De acuerdo
- ○ Muy de acuerdo
- ○ Totalmente de acuerdo

4) La imagen me parece verosímil. *

Escala: 1 = Nada de acuerdo → 7 = Totalmente de acuerdo

- ○ Nada de acuerdo
- ○ Muy en desacuerdo
- ○ En desacuerdo

- ○ Ni de acuerdo ni en desacuerdo
- ○ De acuerdo
- ○ Muy de acuerdo
- ○ Totalmente de acuerdo

5) El contenido de esta imagen no es engañoso. *

Escala: 1 = Nada de acuerdo → 7 = Totalmente de acuerdo

- ○ Nada de acuerdo
- ○ Muy en desacuerdo
- ○ En desacuerdo
- ○ Ni de acuerdo ni en desacuerdo
- ○ De acuerdo
- ○ Muy de acuerdo
- ○ Totalmente de acuerdo

Bloque 2: Autenticidad percibida

Escala de respuesta: 1 = Nada de acuerdo → 7 = Totalmente de acuerdo

6) El contenido de la imagen se siente genuino. *

Escala: 1 = Nada de acuerdo → 7 = Totalmente de acuerdo

- ○ Nada de acuerdo
- ○ Muy en desacuerdo
- ○ En desacuerdo
- ○ Ni de acuerdo ni en desacuerdo

- ○ De acuerdo
- ○ Muy de acuerdo
- ○ Totalmente de acuerdo

7) La imagen me parece natural. *

Escala: 1 = Nada de acuerdo → 7 = Totalmente de acuerdo

- ○ Nada de acuerdo
- ○ Muy en desacuerdo
- ○ En desacuerdo
- ○ Ni de acuerdo ni en desacuerdo
- ○ De acuerdo
- ○ Muy de acuerdo
- ○ Totalmente de acuerdo

8) La persona en la imagen me parece auténtica. *

Escala: 1 = Nada de acuerdo → 7 = Totalmente de acuerdo

- ○ Nada de acuerdo
- ○ Muy en desacuerdo
- ○ En desacuerdo
- ○ Ni de acuerdo ni en desacuerdo
- ○ De acuerdo
- ○ Muy de acuerdo
- ○ Totalmente de acuerdo

9) Siento que la imagen representa una situación realista. *

Escala: 1 = Nada de acuerdo → 7 = Totalmente de acuerdo

- ○ Nada de acuerdo
- ○ Muy en desacuerdo
- ○ En desacuerdo
- ○ Ni de acuerdo ni en desacuerdo
- ○ De acuerdo
- ○ Muy de acuerdo
- ○ Totalmente de acuerdo

10) Percibo el contenido de la imagen como fiel a la realidad. *

Escala: 1 = Nada de acuerdo → 7 = Totalmente de acuerdo

- ○ Nada de acuerdo
- ○ Muy en desacuerdo
- ○ En desacuerdo
- ○ Ni de acuerdo ni en desacuerdo
- ○ De acuerdo
- ○ Muy de acuerdo
- ○ Totalmente de acuerdo

Bloque 3: Confianza en el contenido

Escala de respuesta: 1 = Nada de acuerdo → 7 = Totalmente de acuerdo (excepto ítem 14)

11) Confiaría en una publicación como esta. *

Escala: 1 = Nada de acuerdo → 7 = Totalmente de acuerdo

- Nada de acuerdo
- Muy en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Muy de acuerdo
- Totalmente de acuerdo

12) El contenido me inspira confianza. *

Escala: 1 = Nada de acuerdo → 7 = Totalmente de acuerdo

- Nada de acuerdo
- Muy en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Muy de acuerdo
- Totalmente de acuerdo

13) Me sentiría cómodo/a interactuando con este tipo de publicación (ejemplo: comentar, darle like o compartir). *

Escala: 1 = Nada de acuerdo → 7 = Totalmente de acuerdo

- Nada de acuerdo
- Muy en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo

- ○ De acuerdo
- ○ Muy de acuerdo
- ○ Totalmente de acuerdo

14) ¿Qué tan identificado/a se siente con el estilo del contenido observado? *

Escala: 1 = Nada identificado/a → 7 = Totalmente identificado/a

- ○ Nada identificado/a
- ○ Muy poco identificado/a
- ○ Poco identificado/a
- ○ Ni identificado/a ni no identificado/a
- ○ Identificado/a
- ○ Muy identificado/a
- ○ Totalmente identificado/a

15) Me fiaría de este contenido al momento de formar una opinión. *

Escala: 1 = Nada de acuerdo → 7 = Totalmente de acuerdo

- ○ Nada de acuerdo
- ○ Muy en desacuerdo
- ○ En desacuerdo
- ○ Ni de acuerdo ni en desacuerdo
- ○ De acuerdo
- ○ Muy de acuerdo
- ○ Totalmente de acuerdo

Bloque 4: Verificación de manipulación

Importante: estas preguntas buscan verificar qué percibió usted al ver la imagen.

16) Según su percepción, ¿la imagen observada fue generada con inteligencia artificial? *

- ○ Sí
- ○ No
- ○ No estoy seguro/a

17) ¿Qué tan seguro/a está de su respuesta anterior? *

Escala: 1 = Nada seguro/a → 7 = Totalmente seguro/a

- ○ Nada seguro/a
- ○ Muy poco seguro/a
- ○ Poco seguro/a
- ○ Ni seguro/a ni inseguro/a
- ○ Seguro/a
- ○ Muy seguro/a
- ○ Totalmente seguro/a

18) ¿Qué tan probable cree que la imagen haya sido creada con IA? *

Escala: 1 = Nada probable → 7 = Totalmente probable

- ○ Nada probable
- ○ Muy poco probable
- ○ Poco probable
- ○ Ni probable ni improbable
- ○ Probable
- ○ Muy probable

- Totalmente probable

19) ¿Recuerda haber visto un ícono o etiqueta en la imagen (por ejemplo, una señal que indicaba algo sobre el origen de la imagen)? *

- Sí
- No

Bloque 5: Actitudes hacia el ícono de transparencia (condiciones con ícono)

Nota: este bloque fue respondido únicamente por los participantes asignados a las condiciones experimentales con ícono de “Información de IA”.

20) En general, ¿qué tan de acuerdo está usted con que el contenido publicado en redes sociales advierta sobre el uso de inteligencia artificial mediante un ícono? *

Escala: 1 = Nada de acuerdo → 7 = Totalmente de acuerdo

- Nada de acuerdo
- Muy en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Muy de acuerdo
- Totalmente de acuerdo

21) ¿Qué tan seguro/a está de su respuesta anterior? *

Escala: 1 = Nada seguro/a → 7 = Totalmente seguro/a

- Nada seguro/a
- Muy poco seguro/a

- ○ Poco seguro/a
- ○ Ni seguro/a ni inseguro/a
- ○ Seguro/a
- ○ Muy seguro/a
- ○ Totalmente seguro/a

22) El ícono presente en la imagen influyó en cómo evalué el contenido. *

Escala: 1 = Nada de acuerdo → 7 = Totalmente de acuerdo

- ○ Nada de acuerdo
- ○ Muy en desacuerdo
- ○ En desacuerdo
- ○ Ni de acuerdo ni en desacuerdo
- ○ De acuerdo
- ○ Muy de acuerdo
- ○ Totalmente de acuerdo

23) El ícono me hizo reflexionar sobre el origen del contenido. *

Escala: 1 = Nada de acuerdo → 7 = Totalmente de acuerdo

- ○ Nada de acuerdo
- ○ Muy en desacuerdo
- ○ En desacuerdo
- ○ Ni de acuerdo ni en desacuerdo
- ○ De acuerdo
- ○ Muy de acuerdo
- ○ Totalmente de acuerdo

24) La presencia del ícono afectó mi nivel de confianza en la imagen. *

Escala: 1 = Nada de acuerdo → 7 = Totalmente de acuerdo

- ○ Nada de acuerdo
- ○ Muy en desacuerdo
- ○ En desacuerdo
- ○ Ni de acuerdo ni en desacuerdo
- ○ De acuerdo
- ○ Muy de acuerdo
- ○ Totalmente de acuerdo

25) El ícono me hizo cuestionar la autenticidad del contenido. *

Escala: 1 = Nada de acuerdo → 7 = Totalmente de acuerdo

- ○ Nada de acuerdo
- ○ Muy en desacuerdo
- ○ En desacuerdo
- ○ Ni de acuerdo ni en desacuerdo
- ○ De acuerdo
- ○ Muy de acuerdo
- ○ Totalmente de acuerdo

26) Considero adecuado que este tipo de contenidos incluya un ícono informativo. *

Escala: 1 = Nada de acuerdo → 7 = Totalmente de acuerdo

- ○ Nada de acuerdo
- ○ Muy en desacuerdo

- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Muy de acuerdo
- Totalmente de acuerdo

27) La señalización visual del ícono me pareció clara y comprensible. *

Escala: 1 = Nada de acuerdo → 7 = Totalmente de acuerdo

- Nada de acuerdo
- Muy en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Muy de acuerdo
- Totalmente de acuerdo

A continuación, se le solicitará información general no identificatoria que será utilizada únicamente con fines académicos. La encuesta es anónima y su identidad estará completamente resguardada.

28) Edad *

- Menor de 18 años
- 18 - 21 años
- 22 - 25 años
- 26 - 28 años
- Mayor de 28 años

29) Sexo *

- ○ Mujer
- ○ Hombre
- ○ Otro / Prefiero no responder

30) ¿Dónde reside usted? *

- ○ Región Metropolitana
- ○ Quinta región
- ○ Otro lugar
- ○ Prefiero no responder

31) ¿Cuál es su nivel de educación? Considere solo niveles cursados y aprobados con el certificado correspondiente. *

- ○ Educación media incompleta
- ○ Educación media completa
- ○ Técnico incompleto
- ○ Técnico completo
- ○ Universitario incompleto
- ○ Universitario completo (titulado)
- ○ Postgrado (magíster o doctorado)

32) ¿Con qué frecuencia usted usa redes sociales en un día típico? *

- ○ Nunca
- ○ Menos de una vez por semana
- ○ 1 - 2 veces por semana

- Una vez al día
- Varias veces al día
- Muchas veces al día
- Casi constantemente durante el día

33) En general, ¿qué tan atento/a suele estar al contenido visual que ve en redes sociales? *

- Nada atento/a
- Muy poco atento/a
- Poco atento/a
- Ni atento/a ni desatento/a
- Atento/a
- Bastante atento/a
- Muy atento/a

Comentario final

34) Comentario final (opcional)

Si lo desea, puede agregar cualquier comentario adicional sobre la imagen o el contenido observado.

[Respuesta abierta]

Cierre y explicación del estudio

¡Muchas gracias por tu participación!

Explicación sobre los objetivos del estudio (retroalimentación entregada a los participantes al finalizar el experimento)

Esta investigación tuvo como objetivo analizar cómo ciertos elementos visuales presentes en contenidos difundidos en redes sociales influyen en la percepción que las personas tienen sobre dichos contenidos. En particular, el estudio busca comprender cómo se evalúan imágenes asociadas a influencers digitales en términos de credibilidad, autenticidad y confianza.

Para ello, se utilizaron diferentes condiciones experimentales que variaban ciertos elementos visuales del contenido observado. Estas variaciones permiten analizar de qué manera la presencia o ausencia de señales informativas, como íconos o etiquetas que explicitan "Información de IA", puede influir en la forma en que las personas interpretan y valoran este tipo de publicaciones en entornos digitales.

Si tienes alguna pregunta, comentario o deseas obtener más información sobre el estudio, puedes contactarnos en los siguientes correos:

Matías Vargas — matias.vargas@itatour.cl

Catalina Urzúa — caturzua@alumnos.uai.cl

Prof. Carmina Rodríguez — carmina.rodriguez@uai.cl

Agradecemos sinceramente tu tiempo y tu colaboración en esta investigación.

Anexo 2: Prompt utilizado para generar la imagen con inteligencia artificial ChatGPT

“Recrea la imagen adjunta con el máximo nivel de realismo fotográfico posible, manteniendo de forma absolutamente fiel y precisa cada uno de sus elementos visuales. No modifiques, omitas ni agregues ningún detalle: conserva exactamente la composición, iluminación, perspectiva, proporciones, colores, texturas, fondo, objetos, reflejos, sombras y disposición espacial original. La reproducción debe ser idéntica a la imagen de referencia, respetando íntegramente su estructura visual y estética, con un acabado hiperrealista y de alta calidad”.