¿Por qué a veces una imagen o vídeo generado por inteligencia artificial nos desasosiega aunque parezca casi real? El fenómeno conocido como «valle inquietante» describe precisamente esa respuesta de extrañeza o rechazo que sentimos ante reproducciones casi humanas.

IMAGEN: FOTOGRAMA DEL *VIDEOCLIP* *ORAL* (Björk ft. Rosalía)[1]

¿Por qué a veces una imagen o vídeo generado por inteligencia artificial nos desasosiega aunque parezca casi real? El fenómeno conocido como «valle inquietante» describe precisamente esa respuesta de extrañeza o rechazo que sentimos ante reproducciones casi humanas.

Originalmente formulado en 1970 por el roboticista Masahiro Mori, el valle inquietante plantea que cuanto más se parecen un robot o figura artificial a un ser humano, más positiva es la reacción... hasta que la similitud casi perfecta provoca repulsión.

En los últimos años, con IA generativa capaz de producir rostros e incluso vídeos realistas, esto ha cobrado nueva relevancia: ¿cómo percibimos los humanos estas creaciones sintéticas? ¿Somos capaces de notar que son artificiales? ¿Por qué a veces no logramos detectarlo?

El valle inquietante es una hipótesis que describe la reacción emocional negativa ante entidades artificiales muy humanas, pero no del todo auténticas. Cuando una figura antropomórfica, un robot, un *avatar* digital, un rostro generado por IA, se acerca mucho a la apariencia humana pero muestra algo sutilmente “fuera de lugar”, suele provocarnos desasosiego. Nuestro cerebro percibe que “algo no encaja”, generando inquietud o simple rechazo.

El valle inquietante es una hipótesis que describe la reacción emocional negativa ante entidades artificiales muy humanas, pero no del todo auténticas

Diversas teorías intentan explicar las causas de este efecto: desde razones evolutivas (nuestro cerebro asociaría las distorsiones faciales con enfermedad o peligro, activando una respuesta de aversión instintiva) hasta cognitivas (la incertidumbre de no poder clasificar algo como humano o no humano genera rechazo) y existenciales (un doble artificial casi idéntico a nosotros puede recordarnos nuestra propia mortalidad o reemplazabilidad).

Desde la neurociencia cognitiva, comienzan a hallarse mecanismos cerebrales detrás del valle inquietante. Investigadores de la Universidad de Cambridge mostraron imágenes de humanos reales, rostros virtuales y robots a voluntarios mientras medían su actividad cerebral por fMRI. Encontraron que el cerebro funciona como una especie de «detector de humanidad»: la corteza prefrontal ventromedial aumentaba su actividad ante figuras más humanizadas pero caía abruptamente al rozar el límite de lo humano sin serlo, mientras que la amígdala se activaba con intensidad, sugiriendo una respuesta emocional de alarma.

Los seres humanos somos expertos en rostros y en descifrar señales sociales sutiles; desde bebés aprendemos a leer expresiones, seguir miradas y distinguir individuos. Esta maestría perceptiva explica por qué podemos notar detalles ínfimos fuera de lugar en una imagen de rostro humano. Ante fotografías o vídeos generados por IA, muchos usuarios reportan que “hay algo en la mirada” o “una sensación rara” que les

delata que no son reales.

Ante fotografías o vídeos generados por IA, muchos usuarios reportan que “hay algo en la mirada” o “una sensación rara” que les delata que no son reales

Hasta hace poco, las imágenes sintéticas solían delatarse por fallos evidentes: manos con seis dedos, ojos asimétricos, texturas de piel irreales. Pero incluso sin errores obvios, nuestro cerebro capta algo: una mirada sin brillo, un gesto congelado, una falta de sincronía entre apariencia y «vida» interior.

Los datos recientes son elocuentes. Las imágenes faciales generadas con ChatGPT y DALL·E resultan virtualmente indistinguibles de fotografías auténticas para la mayoría de observadores. Los programas de IA alcanzan un 97 % de precisión detectando rostros sintéticos en fotos, pero los humanos no superamos el porcentaje atribuible al azar; curiosamente, con vídeos *deepfake* la situación se invertía y los humanos acertaban dos tercios de las veces. Incluso los «súper-reconocedores», el 2 % superior en reconocimiento facial, apenas detectan el 41 % de los rostros falsos, una tasa inferior al azar.

Sin embargo, cinco minutos de entrenamiento sobre errores comunes de renderizado mejoró sustancialmente su precisión. Es decir, que nuestro sistema perceptivo no está calibrado para esta amenaza, pero puede entrenarse.

Frente el rápido progreso de la inteligencia artificial generativa, surge la pregunta: ¿podrán las máquinas cruzar el valle inquietante, eliminando esa inquietud por completo? Los avances recientes apuntan en esa dirección.

En el campo de las imágenes estáticas, los generadores basados en redes antagónicas generativas (GAN) y modelos de difusión han logrado crear rostros y cuerpos virtuales indistinguibles de fotografías reales. Las caras generadas por StyleGAN2 ya alcanzan un nivel de detalle anatómico y calidad fotográfica que engaña a la mayoría de observadores.

Lo estamos viendo también en ejemplos cotidianos. Los verificadores de contenido ahora hablan de una “perfección inquietante” como nueva señal: fotogramas con personas de belleza impecable, sin ninguna arruga fuera de lugar, con simetrías casi matemáticas.

Paradójicamente, la IA crea imágenes tan pulidas que producen otra forma de artificio: no por defectos grotescos, sino por ausencia de las pequeñas imperfecciones que dan autenticidad. Aun así, para la mayoría del público esas minucias pasan inadvertidas.

El desafío mayor, sin embargo, está en el vídeo. No basta con un fotograma realista; hay que encadenar miles por segundo sin caer en gestos espasmódicos o inexpresivos. Hasta hace poco, los primeros sistemas de texto a vídeo producían resultados entre lo cómico y lo espeluznante: clips borrosos, figuras humanas inestables que parecían salidas de un sueño raro... Pero la velocidad con la que esto está cambiando resulta difícil de exagerar.

En los últimos meses se han sucedido lanzamientos que redefinen lo posible. Google DeepMind presentó Veo 3.1 en octubre de 2025, un modelo que trata el sonido como parte integral del vídeo: genera diálogos con labios sincronizados, efectos de sonido alineados con la acción y paisajes sonoros ambientales. No es un detalle menor: una de las pistas clásicas para detectar un vídeo falso era la desincronización entre labios y voz. Cuando eso desaparece, una barrera perceptiva cae con ello.

Una de las pistas clásicas para detectar un vídeo falso era la desincronización entre labios y voz

En febrero de 2026, la empresa china Kuaishou lanzó Kling 3.0, que permite generar hasta seis tomas distintas dentro de un mismo clip de 15 segundos manteniendo la coherencia de personajes y escenarios, con resolución 4K y sincronización labial en múltiples idiomas. Lo que importa para el valle inquietante es la consistencia temporal: cuando cada fotograma se genera teniendo en cuenta decenas de fotogramas adyacentes, las «mutaciones» faciales que antes delataban el origen artificial se reducen drásticamente.

Pero el modelo que más debate ha generado es Seedance 2.0, de ByteDance. Clips virales mostraron a Brad Pitt y Tom Cruise en una pelea coreografiada tan convincente que Disney envió una carta de cese y desistimiento y Paramount acusó a la compañía de infracción de propiedad intelectual.

¿Se ha cruzado entonces el valle? No del todo. Los modelos de 2026 todavía luchan con acciones cotidianas: comer, manipular cubiertos, interactuar con objetos pequeños. No tenemos referencia de cómo se mueve un dragón, pero hemos visto a miles de personas comer pasta, y cualquier desviación salta a la vista. A esto se suma que los modelos de imagen estática, como la familia Nano Banana de Google, ya sirven como fotogramas de referencia para los generadores de vídeo, minimizando las incoherencias entre cuadros que antes delataban el contenido sintético.

Un último dato que ayuda a enmarcar la velocidad del cambio: el número de *deepfakes* en internet pasó de unos 500 000 en 2023 a unos 8 millones en 2025, con un crecimiento anual cercano al 900 %. Un investigador de la Universidad de Buffalo especializado en medios sintéticos escribió en Fortune que la clonación de voz ha cruzado lo que él llama el “umbral de la indistinguibilidad”: unos pocos segundos de audio bastan para generar un clon convincente con entonación, ritmo, pausas y hasta ruido de respiración naturales.

No parece que el valle inquietante se limite a lo visual: también se manifiesta en interacciones textuales con *chatbots*. Sin embargo, los usuarios siguen prefiriendo la naturalidad y las imperfecciones humanas: mientras que los defectos humanos aumentan la cercanía, las desviaciones que rompen la percepción de humanidad disparan el rechazo.

Los usuarios siguen prefiriendo la naturalidad y las imperfecciones humanas

El sector tecnológico está respondiendo con soluciones de verificación que funcionan donde nuestros ojos ya no pueden. La lógica es sencilla: si no podemos ver la diferencia, al menos podemos marcar el contenido en el momento de su creación. Estas marcas sobreviven a compresiones, recortes y conversiones de formato habituales.

Desde mayo de 2025, un portal de verificación de Google DeepMind permite comprobar si un archivo contiene SynthID, una marca de agua imperceptible que se inserta durante la generación. En paralelo, la C2PA (Coalition for Content Provenance and Authenticity), impulsada por Adobe, Microsoft, Google, OpenAI y Meta, desarrolla un estándar abierto que adjunta al archivo información criptográfica verificable sobre su origen y ediciones. Mientras SynthID es la huella invisible que persiste cuando se pierde el metadato, C2PA ofrece la trazabilidad cuando las plataformas lo preservan.

La regulación también avanza, aunque fragmentada. El Reglamento de IA de la Unión Europea, en vigor desde agosto de 2024, exige que todo contenido generado por IA sea marcado en formato legible por máquinas, con pleno cumplimiento requerido para agosto de 2026. Pero el panorama industrial muestra a cada gran empresa desarrollando su propio sistema, sin un estándar universal de detección.

La percepción del valle inquietante es un fascinante cruce entre biología, mente y tecnología. Sentimos inquietud ante lo casi humano porque nuestros cerebros están calibrados finamente para reconocer a nuestros semejantes y detectar lo que se aparta de la norma. Esa misma agudeza se activa con las creaciones de IA que casi logran imitarnos.

A comienzos de 2026, el estado de la cuestión es claro: la frontera se desplaza a una velocidad vertiginosa. Lo que en enero de 2026 era limitación de un modelo, en febrero ya lo resolvía el siguiente. Quizás el cambio más profundo no sea visual sino conceptual: en lugar de detectar “algo extraño”, empezaremos a desconfiar de “algo demasiado perfecto”.

Quizás el cambio más profundo no sea visual sino conceptual: en lugar de detectar algo ‘extraño’, empezaremos a desconfiar de ‘algo demasiado perfecto’

¿Desaparecerá por completo el valle inquietante? Probablemente no: seguiremos teniendo reparo ante un robot físico que intenta ser nuestro doble perfecto. Pero en el terreno visual digital, la distinción entre lo generado y lo real dependerá cada vez más de la tecnología que nos asiste. Cuando ya no podamos confiar en «lo noto en mi estómago, se ve falsa», necesitaremos marcas de agua universales, credenciales de procedencia y, sobre todo, educación mediática para orientarnos en un mundo donde lo artificial se camufla con total naturalidad.