LA VOZ EN LA ENCRUCIJADA DIGITAL: IMPLICACIONES ÉTICAS Y AXIOLÓGICAS ANTE LA TECNOLOGÍA DE CLONACIÓN Y EL DEEPFAKE

AUTOR:

DÍAZ, JOSUE

RESUMEN

La voz humana constituye un elemento central de la comunicación, en tanto expresión de identidad, presencia y vínculo social. Sin embargo, el desarrollo acelerado de tecnologías basadas en inteligencia artificial ha introducido sistemas capaces de clonar, manipular y automatizar la producción y la escucha vocal, generando una transformación profunda de su estatuto comunicacional. Este artículo tiene como objetivo analizar las implicaciones éticas y axiológicas de la clonación de voz, los deepfakes de audio y el reconocimiento automático del habla, situando la voz en el contexto de una encrucijada digital. Desde un enfoque cualitativo, se realizó una revisión documental con perspectiva hermenéutica, integrando aportes de las Ciencias de la Comunicación, la filosofía de la técnica y los estudios sobre inteligencia artificial. El análisis se apoya en el transhumanismo como marco filosófico, problematizado a la luz del pensamiento de Martin Heidegger, para examinar la reducción técnica de la voz como recurso disponible. Se concluye que estas tecnologías comprometen la autenticidad, la confianza y el valor axiológico de la voz, planteando la necesidad de una reflexión ética orientada a preservar su dimensión humana en la comunicación contemporánea.

Palabras clave: voz humana; inteligencia artificial; clonación de voz; deepfakes de audio; ética de la comunicación; transhumanismo.

ABSTRACT

The human voice constitutes a central element of communication, as it embodies identity, presence, and social connection. However, the accelerated development of artificial intelligence–based technologies has enabled systems capable of cloning, manipulating, and automating both the production and the perception of voice, profoundly transforming its communicational status. This article aims to analyze the ethical and axiological implications of voice cloning, audio deepfakes, and automatic speech recognition, situating the voice within a contemporary digital crossroads. Using a qualitative approach, a documentary review with a hermeneutic perspective was conducted, integrating contributions from Communication Sciences, philosophy of technology, and artificial intelligence studies. The analysis adopts transhumanism as a philosophical framework, critically examined through Martin Heidegger's reflections on technology, to problematize the technical reduction of the voice as a disposable resource. The findings suggest that these technologies undermine authenticity, trust, and the axiological value of the voice, highlighting the need for ethical reflection aimed at preserving its human dimension in contemporary communication.

Keywords: human voice; artificial intelligence; voice cloning; audio deepfakes; communication ethics; transhumanism.

Introducción

En las sociedades contemporáneas, la voz humana ocupa un lugar central en los procesos comunicacionales, al constituirse como uno de los principales vehículos de expresión, interacción social y construcción de confianza. A través de la voz, los individuos no solo transmiten información, sino que manifiestan identidad, intencionalidad y presencia, elementos fundamentales para la comprensión mutua y la cohesión social. Esta condición ha otorgado históricamente a la voz un valor simbólico y axiológico que trasciende su dimensión sonora, consolidándola como un componente esencial de la comunicación humana.

No obstante, el desarrollo acelerado de tecnologías basadas en inteligencia artificial ha comenzado a transformar de manera significativa la relación tradicional entre voz, identidad y autenticidad en los procesos comunicacionales. En particular, la emergencia de sistemas capaces de simular, clonar y manipular voces humanas con altos niveles de realismo ha generado un escenario inédito en el que la voz puede ser separada de su emisor original. Según Javier Franganillo, "La inteligencia artificial (IA) generativa es un campo en rápido avance que ya permite la producción automatizada de contenido textual, gráfico, sonoro y audiovisual de alta calidad".(Franganillo, 2023, p. 1), esto plantea desafíos significativos en relación con la autenticidad de los contenidos digitales, al posibilitar la producción de materiales que pueden simular la autoría humana sin la intervención directa del sujeto, lo que a su vez genera interrogantes profundos sobre el estatuto comunicacional de la voz y su fiabilidad como indicador de presencia y autoría.

En los últimos años, el desarrollo de sistemas de inteligencia artificial capaces de generar contenidos audiovisuales sintéticos ha dado lugar a fenómenos conocidos como deepfakes, los cuales plantean importantes desafíos para la seguridad, la privacidad y la integridad de la información en

entornos digitales. Diversos estudios advierten que estas tecnologías pueden utilizarse para manipular identidades, difundir información falsa o cometer fraudes mediante suplantación, lo que introduce nuevas formas de daño reputacional, emocional y social. En este sentido, se ha señalado que “Los deepfakes también podrían utilizarse para retratar falsamente a alguien como si estuviera respaldando un producto, un servicio, una idea o un político.” (Chesney & Citron, 2019, p. 1774).

En este punto, la generación artificial de voz ocupa un lugar particularmente sensible dentro del ecosistema de los deepfakes, ya que la voz constituye uno de los principales marcadores de identidad y autenticidad en la comunicación humana. En este sentido, “la IA también se está utilizando para sintetizar audio de alta calidad que imita voces humanas” (Vaccari y Chadwick, 2020, p. 3). A partir de ello, la posibilidad de replicar o sintetizar voces con alto grado de verosimilitud amplía el potencial de estas tecnologías para la manipulación comunicativa y la suplantación de identidad, generando escenarios que desafían los marcos tradicionales de verificación de la información.

A nivel institucional, organismos internacionales han advertido sobre el uso creciente de contenidos generados por inteligencia artificial en contextos delictivos y de desinformación. Informes recientes señalan que "muchas organizaciones han comenzado ahora a considerar los deepfakes como un riesgo potencial aún mayor que el robo de identidad" (Europol, 2023, p. 4). En este contexto, los deepfakes facilitan fraudes y manipulación al aumentar la credibilidad de contenidos falsificados. Asimismo, se concluye que "es altamente probable que los actores maliciosos hagan un uso creciente de la tecnología deepfake para facilitar diversos actos delictivos y llevar a cabo campañas de desinformación con el fin de influir en la opinión pública o distorsionarla" (Europol, 2023, p. 22). En consecuencia, este fenómeno constituye un desafío tecnológico, social y comunicativo que exige nuevas herramientas analíticas.

A estas tensiones se suma un debate de carácter filosófico más amplio, vinculado al transhumanismo y a la progresiva sustitución de capacidades humanas por sistemas tecnológicos. Desde esta perspectiva, la intervención técnica sobre la voz no puede comprenderse únicamente como un avance instrumental, sino como parte de un proceso que reconfigura las relaciones entre cuerpo, tecnología y comunicación. En este sentido, la voz se sitúa hoy en una auténtica encrucijada digital, donde convergen innovación tecnológica, vacíos éticos y desafíos axiológicos.

En este contexto, el presente artículo tiene como objetivo analizar críticamente las implicaciones éticas y axiológicas de la clonación de voz y los deepfakes de audio, partiendo de la voz humana como categoría central de la comunicación. A través de una revisión documental con enfoque hermenéutico, se propone reflexionar sobre los efectos de estas tecnologías en la autenticidad comunicacional, la identidad y la confianza social, aportando elementos teóricos para comprender y problematizar el lugar de la voz en la comunicación contemporánea.

Marco teórico:

La voz como fenómeno humano, comunicacional y social

La voz humana constituye un fenómeno complejo que articula dimensiones biológicas, psicológicas, sociales y simbólicas, lo que la convierte en un elemento central de los procesos comunicacionales. Desde el campo de la vocología y la fonoaudiología, Ingo Titze considerado por muchos el padre de la vocología plantea que la voz puede entenderse como un sistema "compuesto principalmente por los pulmones, la tráquea, la glotis y el tracto respiratorio superior, situado entre la glotis y la boca. Este tracto, llamado también tracto vocal, desempeña un papel clave en la formación de los sonidos vocales." (Titze, 2000, p. 65). Esta definición, ampliamente aceptada en la literatura especializada, permite comprender la voz no como un simple producto acústico, sino como un proceso fisiológico altamente especializado.

En el contexto del presente artículo, esta concepción resulta clave, ya que evidencia que toda manifestación vocal está necesariamente vinculada a un cuerpo, a un sistema biológico y a una experiencia humana situada.

Desde una perspectiva comunicacional, la voz trasciende su dimensión fisiológica para convertirse en un vehículo de sentido, identidad y relación social. Según Poyatos (2002), la voz funciona como un marcador paralingüístico, tomando en cuenta la "frecuencia, intensidad y duración, podemos inferir ciertos rasgos de la personalidad y, con mayor precisión, el estado de ánimo, el estado emocional y el estado de salud" (Poyatos, 2002. p.267). Este planteamiento permite afirmar que la voz no solo comunica lo que se dice, sino también quién lo dice y desde dónde se dice, reforzando su papel como elemento constitutivo de la identidad comunicativa del sujeto. En conexión con este artículo, esta característica explica por qué la voz ha sido históricamente asociada a la autenticidad, la presencia y la confianza en las interacciones humanas.

En el ámbito social, la voz desempeña un papel fundamental en la construcción de ciudadanía y en la participación en la vida pública. Couldry (2010), sostiene que "tener voz exige recursos tanto prácticos (el lenguaje) como de estatus (aparentemente simbólico) que resultan indispensables para ser reconocido por los demás como alguien que posee voz... Por consiguiente, una concepción no social (o puramente individual) de la voz pasaría por alto una dimensión vital" (Couldry, 2010, p. 8). A partir de esta perspectiva, se amplía el concepto de voz más allá de sus dimensiones biológicas o técnicas, entendiéndola como un derecho comunicacional y un recurso de poder simbólico. En consecuencia, la voz se configura como un elemento estructurante de las relaciones sociales, capaz de generar transformaciones significativas en la vida de las personas a través del propio acto de comunicarse.

Desde esta perspectiva integral, la voz puede entenderse como un fenómeno sistémico en el que convergen cuerpo, subjetividad y

sociedad. Esta comprensión resulta especialmente relevante en el contexto tecnológico contemporáneo, donde el desarrollo de herramientas basadas en inteligencia artificial comienza a intervenir, reproducir y simular la voz humana. La posibilidad de separar la voz de su fuente corporal y de su emisor original introduce una ruptura profunda en la relación tradicional entre voz, identidad y responsabilidad comunicativa. Es precisamente en este punto donde emerge la preocupación central de este artículo, al evidenciarse que la naturalidad, la originalidad y el valor axiológico de la voz se ven tensionados por tecnologías que cuestionan su condición de expresión genuinamente humana.

Así, comprender qué es la voz, por qué es importante y cómo impacta en las relaciones humanas constituye el punto de partida indispensable para analizar críticamente los fenómenos de clonación vocal y deepfakes de audio. Solo desde este reconocimiento de la voz como bien comunicacional y humano es posible dimensionar la magnitud de la encrucijada digital que plantea su intervención tecnológica, y avanzar hacia una reflexión ética y axiológica acorde con los desafíos actuales de las Ciencias de la Comunicación.

La voz en la comunicación contemporánea

En la comunicación contemporánea, la voz mantiene un papel central como soporte de sentido, credibilidad y vínculo social, incluso en entornos mediados por tecnologías digitales. Diversos estudios han destacado que funciona como un marcador de autenticidad, permitiendo a los interlocutores evaluar la veracidad, la intención y la fiabilidad del mensaje. En este marco, las cualidades vocales —como el tono, el timbre, el ritmo y la prosodia— se constituyen en recursos semióticos. Como señala Van Leeuwen (1999), “la semiótica del sonido se ocupa de describir qué se puede decir mediante el sonido y cómo se pueden interpretar las cosas que otras personas dicen con sonido” (Van Leeuwen, 1999, p. 4). De este modo, la voz no se limita a ser un canal técnico, sino que actúa como

un componente activo en la producción de sentido comunicacional, aspecto clave para el análisis desarrollado en este artículo.

En entornos mediados, como la radio, el podcast, las plataformas de mensajería de voz y los sistemas de atención automatizada, la voz ldry (2012) señala que las representaciones mediadas se inscriben en prácticas sociales más amplias que nos sitúan en el mundo y nos conectan con otros, ofreciendo "el potencial de reconocimiento mutuo." (Couldry, 2012, p. 181). Esta perspectiva resulta especialmente relevante para el presente trabajo, ya que evidencia que la centralidad de la voz no ha disminuido con la digitalización, sino que se ha reconfigurado, ampliando su alcance y su impacto en la esfera pública y privada.

No obstante, esta centralidad de la voz en la comunicación se ve tensionada por los procesos de automatización y mediatización propios de la cultura digital. Como señalan Couldry y Hepp (2017), en las sociedades profundamente mediadas las formas tradicionales de continúa siendo un elemento privilegiado de interacción. Incluso en ecosistemas digitales altamente tecnificados, la voz conserva su capacidad de generar reconocimiento social y de establecer relaciones de cercanía simbólica entre emisores y receptores. En este sentido, Cou interacción se transforman, dando lugar a entornos en los que los actores humanos comparten el espacio comunicacional con sistemas técnicos capaces de producir y reproducir mensajes. En este sentido, "Estas aplicaciones se nos presentan como interlocutores cuasihumanos con los que (mediante software de reconocimiento de voz) podemos comunicarnos utilizando nuestra voz natural" (Couldry y Hepp, 2017, p. 188). Así, la voz comienza a ser gestionada, modulada y reproducida por tecnologías que alteran la relación directa entre el sujeto hablante y su expresión vocal, marcando un punto de inflexión que se vincula con la problemática de la clonación de voz abordada posteriormente.

Desde una perspectiva ética, la voz en la comunicación contemporánea

se encuentra estrechamente vinculada a la noción de confianza. Luhmann (2005) afirma que “la confianza es un mecanismo de reducción de la complejidad social” (Luhmann, 2005, p XXII), permitiendo a los individuos interactuar sin verificar constantemente la autenticidad de cada mensaje. Aplicado al ámbito vocal, este planteamiento permite entender por qué la voz ha sido tradicionalmente percibida como una señal fiable de presencia y autoría. En conexión con este artículo, la posibilidad de simular voces humanas mediante inteligencia artificial amenaza este principio, al introducir incertidumbre sobre la procedencia y legitimidad del mensaje vocal.

Asimismo, en el campo de la comunicación pública y política, la voz desempeña un papel determinante en la construcción de liderazgo, legitimidad y persuasión. Desde el enfoque del análisis crítico del discurso, Fairclough (2010) sugiere "que existen dos relaciones entre el discurso —y la realidad más allá del discurso— asociadas a la convicción: la construcción del consenso y la construcción de la legitimidad" (Fairclough, 2010, p. 536), lo que refuerza su importancia en la configuración de imaginarios sociales y relaciones de poder. Esta dimensión adquiere especial relevancia en un escenario donde las voces pueden ser artificialmente replicadas, ya que la manipulación vocal no solo afecta a individuos concretos, sino que tiene el potencial de incidir en procesos colectivos de toma de decisiones.

En síntesis, la voz en la comunicación contemporánea se sitúa en una posición ambivalente: por un lado, conserva su valor como recurso comunicacional esencial para la construcción de sentido, confianza y vínculo social; por otro, se ve expuesta a procesos tecnológicos que cuestionan su autenticidad y su anclaje en el sujeto humano. Esta tensión constituye el sustrato conceptual que permite comprender la emergencia de tecnologías como la clonación de voz y los deepfakes de audio no como fenómenos aislados, sino como manifestaciones de una transformación más amplia de la comunicación en la era digital,

transformación que será analizada a continuación.

Clonación de voz: intervención tecnológica e implicaciones éticas

La clonación de voz se ha consolidado en los últimos años como una de las aplicaciones más avanzadas de la inteligencia artificial en el ámbito sonoro, al permitir la reproducción sintética de voces humanas con elevados niveles de fidelidad acústica. Desde una perspectiva técnica, esta tecnología se apoya en modelos de aprendizaje profundo capaces de analizar grandes volúmenes de datos vocales para capturar patrones asociados al timbre, la prosodia y la entonación de un hablante específico. En los sistemas contemporáneos de reconocimiento y modelado de voz, el análisis del habla se basa en la extracción de características acústicas y prosódicas; en este sentido, se señala que “los rasgos prosódicos y espectro-temporales incluyen el tono, la energía, el ritmo, la duración y los atributos temporales,” permiten representar matemáticamente las particularidades de la voz de cada individuo. (Sukumar et al., 2024, p. 338). Esta capacidad técnica constituye el punto de partida para comprender la magnitud del fenómeno abordado en este artículo.

No obstante, más allá de su funcionamiento tecnológico, la clonación de voz plantea implicaciones comunicacionales profundas. Al reproducir una voz sin la participación del sujeto que la encarna, se rompe la relación histórica entre emisión vocal, presencia corporal y responsabilidad comunicativa. En este sentido, Chesney y Citron (2019) advierten que la capacidad de distorsionar la realidad ha dado un salto exponencial gracias a la tecnología deep fake, ya que permiten crear representaciones altamente realistas de personas “diciendo o haciendo cosas que nunca dijeron o hicieron” (Chesney y Citron, 2019, p. 1753). Esta posibilidad introduce una ruptura en los criterios tradicionales de autoría y agencia comunicativa resultando central para el presente análisis, ya que permite comprender la clonación

vocal no solo como una innovación técnica, sino como un fenómeno que altera las bases simbólicas de la comunicación interpersonal.

Desde una perspectiva ética, diversos estudios advierten que la clonación de voz transforma la identidad vocal en un recurso susceptible de apropiación indebida. Las tecnologías de síntesis vocal pueden ser explotadas para fraudes, suplantación de identidad y ataques de ingeniería social basados en la confianza interpersonal (Europol, 2023, pp. 9–10). Esta problemática conecta directamente con el eje central del artículo, en tanto evidencia que la voz deja de ser una expresión personal e intransferible para convertirse en un objeto técnico replicable, descontextualizado de la experiencia humana que le da origen.

En el ámbito social, la clonación de voz ha mostrado un impacto creciente en delitos asociados a la suplantación de identidad y el fraude comunicacional. Diversos informes documentan su uso en estafas telefónicas y estrategias de ingeniería social basadas en la confianza en la comunicación oral. Puede citarse el caso “en el que unos criminales utilizaron audio generado mediante deepfake para suplantar la identidad del director ejecutivo de una empresa, con el fin de lograr que un empleado transfiriera 35 millones de dólares” (Europol, 2023, p. 11). Esto evidencia cómo el audio sintético incrementa la eficacia de los fraudes al apoyarse en la credibilidad de la voz humana. En consecuencia, la clonación vocal debe analizarse como un fenómeno ya presente en la dinámica comunicacional contemporánea.

Desde el campo de las Ciencias de la Comunicación, la clonación de voz obliga a repensar el estatuto de la voz como signo comunicativo. Si la voz puede ser replicada y manipulada de manera independiente del sujeto, se debilita su función como garante de autenticidad y presencia. En este sentido, la clonación vocal inaugura una zona de ambigüedad comunicacional en la que resulta cada vez más complejo distinguir entre lo humano y lo artificial. Esta ambigüedad constituye un componente clave de la encrucijada digital que atraviesa la voz en la

actualidad y prepara el terreno para el análisis de los deepfakes de audio, donde la manipulación vocal alcanza una escala aún mayor.

En definitiva, la clonación de voz representa un punto de inflexión en la relación entre tecnología y comunicación. Si bien ofrece aplicaciones potencialmente legítimas en ámbitos como la accesibilidad o la preservación vocal, su uso indiscriminado y no regulado plantea riesgos éticos y axiológicos significativos. Comprender estos riesgos desde una perspectiva comunicacional resulta indispensable para abordar críticamente las transformaciones que la inteligencia artificial introduce en la voz humana y para avanzar hacia una reflexión más amplia sobre los límites y responsabilidades del desarrollo tecnológico.

Deepfakes de audio y crisis de credibilidad comunicativa

Los deepfakes de audio constituyen una evolución significativa de las tecnologías de síntesis y manipulación vocal, al combinar la clonación de voz con sistemas avanzados de generación de discurso capaces de producir mensajes completos atribuidos falsamente a un hablante determinado. A diferencia de la clonación vocal aislada, los deepfakes de audio no solo reproducen las características acústicas de una voz, sino que permiten construir narrativas sonoras verosímiles, insertas en contextos comunicacionales específicos. En este sentido, se ha señalado que los deepfakes de audio emplean métodos avanzados de inteligencia artificial, aprendizaje automático y redes neuronales “para manipular o sintetizar contenido de audio engañoso” (Zhang et al., 2025, p. 2). Esta capacidad de generar audios altamente realistas y convincentes amplía significativamente el potencial de manipulación comunicacional asociado a estas tecnologías.

Desde una perspectiva comunicacional, los deepfakes de audio amplifican la ruptura entre voz, identidad y autoría previamente introducida por la clonación vocal. Mientras que la voz humana ha funcionado históricamente como un

indicador confiable de presencia y responsabilidad discursiva, la posibilidad de generar discursos vocales falsos socava esta función simbólica. En este sentido, Chesney y Citron, (2019), señalan que los deepfakes permiten producir representaciones mediáticas en las que las personas "aparecen diciendo o haciendo cosas que nunca dijeron o hicieron"(Chesney y Citron, 2019, p. 1760), lo que dificulta distinguir entre contenido auténtico y manipulado. Esta situación contribuye a erosionar la confianza pública en los mensajes difundidos a través de medios digitales y plataformas de comunicación.

En el ámbito social, los deepfakes de audio han comenzado a utilizarse con fines de manipulación informativa, desinformación y fraude, afectando tanto a individuos como a instituciones. Estudios recientes advierten que estos contenidos pueden ser empleados para influir en decisiones personales, económicas y políticas, al explotar la confianza depositada en la voz como señal de legitimidad. Según Chesney y Citron (2019), “Además de infligir daño psicológico directo a las víctimas, la tecnología deepfake puede utilizarse para perjudicarlas en otras dimensiones, debido a su utilidad para el sabotaje reputacional”, lo cual se explica precisamente porque estas tecnologías operan sobre la base de una simulación altamente convincente de la identidad vocal (Chesney & Citron, 2019, p. 1774).

Desde una perspectiva ética, los deepfakes de audio plantean interrogantes profundas sobre la responsabilidad moral de quienes desarrollan, difunden y consumen este tipo de tecnologías. La manipulación deliberada de la voz para inducir a error implica una vulneración de principios fundamentales como la veracidad, el consentimiento y el respeto a la dignidad del sujeto. En este sentido, de Ruiter (1021) señala que los deepfakes resultan moralmente problemáticos cuando utilizan datos digitales de la imagen o la voz de una persona para representarla de formas en las que no querría ser presentada, corriendo "el riesgo de socavar la autonomía de las personas y su capacidad para emprender acciones

acordes con su propia voluntad" (de Ruiter, 2021, p. 10). De este modo, la capacidad de simular la voz humana obliga a repensar los marcos éticos de la comunicación, ya que compromete la relación entre mensaje, emisor y verdad.

En el campo de las Ciencias de la Comunicación, los deepfakes de audio pueden ser interpretados como una manifestación extrema de la crisis de autenticidad sonora. Al posibilitar la producción masiva de discursos vocales falsos, estas tecnologías no solo afectan la comunicación interpersonal, sino que también inciden en la esfera pública, los medios de comunicación y la construcción de opinión. Esta situación refuerza la idea de que la problemática de los deepfakes no es meramente técnica, sino profundamente comunicacional y axiológica.

En pocas palabras, los deepfakes de audio intensifican los desafíos éticos y comunicacionales asociados a la intervención tecnológica sobre la voz humana. Al disolver los límites entre lo auténtico y lo artificial, estos sistemas profundizan la encrucijada digital en la que se sitúa la voz en la actualidad. Comprender su impacto resulta indispensable para analizar de manera crítica las transformaciones contemporáneas de la comunicación y para avanzar hacia marcos de reflexión y regulación que protejan el valor humano y social de la voz.

El transhumanismo como marco filosófico

El transhumanismo se configura como una corriente filosófica y cultural que promueve la superación de las limitaciones biológicas humanas mediante el uso intensivo de la ciencia y la tecnología. El término fue introducido por Julian Huxley en 1957, quien planteó la posibilidad de que la humanidad pudiera trascender su estado actual a través de una evolución consciente y dirigida, apoyada en el desarrollo tecnológico (Huxley, 1957). Esta propuesta sentó las bases de un pensamiento que, en las últimas décadas, ha cobrado especial relevancia en los debates contemporáneos sobre inteligencia artificial, biotecnología y

transformación de la condición humana.

En el transhumanismo contemporáneo, autores como Bostrom sostienen que el desarrollo tecnológico debe entenderse como una vía legítima para ampliar las capacidades humanas. En este sentido, el autor afirma que “Los transhumanistas defienden el derecho moral de aquellos que así lo deseen a utilizar la tecnología para ampliar sus capacidades mentales y físicas, y para mejorar el control sobre sus propias vidas.” (Bostrom, 2005, p. 21). Desde esta perspectiva, se proyecta la posibilidad de un estadio poshumano caracterizado por funciones cognitivas, corporales y comunicacionales potenciadas o externalizadas mediante sistemas artificiales. En el marco del presente artículo, esta visión resulta fundamental para comprender el trasfondo filosófico que impulsa el desarrollo de tecnologías orientadas a la simulación, reproducción y automatización de la voz humana.

Aplicado al ámbito comunicacional, el transhumanismo introduce una lógica de optimización funcional que tiende a concebir la voz como una capacidad técnica susceptible de ser replicada, mejorada o sustituida. La voz, entendida tradicionalmente como un proceso fisiológico y expresivo que articula respiración, fonación, emoción e intención comunicativa, comienza a ser tratada como un conjunto de datos analizables y reproducibles. Desde esta lógica, tecnologías como la clonación de voz y los deepfakes de audio no aparecen como anomalías, sino como expresiones coherentes de un proyecto filosófico que busca integrar progresivamente lo humano con lo artificial.

Desde una perspectiva crítica, el pensamiento de Martin Heidegger ofrece herramientas conceptuales clave para problematizar este proyecto. En su reflexión sobre la técnica moderna, Heidegger (1954/1994) sostiene que la esencia de la técnica no se reduce a los dispositivos tecnológicos, sino que remite a un modo particular de desocultamiento del ser, que denomina Gestell (armazón o encuadramiento). En este marco, advierte que tanto la naturaleza como

el ser humano pueden ser comprendidos como Bestand (fondo disponible), es decir, como recursos susceptibles de ser calculados, gestionados y explotados. Esta concepción resulta especialmente pertinente para analizar el tratamiento tecnológico de la voz en el contexto transhumanista.

Desde la óptica heideggeriana, la voz humana corre el riesgo de ser reducida a un objeto técnico manipulable, despojado de su carácter originario como expresión del Dasein. La clonación vocal y los deepfakes de audio pueden interpretarse, así, como manifestaciones de un proceso de instrumentalización en el que la voz deja de ser comprendida como acontecimiento relacional y se transforma en un recurso funcional sometido a la lógica de la eficiencia y el control. Esta reducción técnica entra en tensión con la concepción de Heidegger del lenguaje como “la casa del ser”, en la medida en que la voz, como expresión viva del lenguaje, es sustituida por simulaciones que carecen de arraigo existencial y responsabilidad ontológica.

En este sentido, el transhumanismo, al privilegiar la optimización tecnológica sobre la experiencia humana, contribuye a una comprensión empobrecida de la voz como fenómeno comunicacional. La externalización de la voz mediante sistemas artificiales no solo afecta su dimensión técnica, sino que compromete su valor axiológico, al desvincularla del cuerpo, de la identidad y de la presencia del hablante. Esta lógica instrumental pone en riesgo la función de la voz como garante de autenticidad, sentido y responsabilidad en la comunicación humana.

En definitiva, el transhumanismo constituye un marco filosófico imprescindible para comprender las transformaciones contemporáneas de la voz en contextos tecnológicos avanzados, pero también plantea desafíos éticos y ontológicos de gran profundidad. A la luz del pensamiento de Heidegger, se hace evidente la necesidad de cuestionar críticamente un desarrollo tecnológico que tiende a reducir la voz a un recurso disponible, olvidando su condición de expresión fundamental del ser-en-el-mundo.

Esta reflexión permite avanzar hacia una comprensión más equilibrada de la relación entre tecnología y comunicación, en la que el progreso técnico no se imponga sobre los valores humanos que sostienen la experiencia comunicacional.

La encrucijada digital de la voz

Las transformaciones tecnológicas analizadas a lo largo de este marco teórico permiten afirmar que la voz humana se encuentra actualmente situada en una auténtica encrucijada digital. Este concepto alude a un punto de tensión en el que convergen, por un lado, el desarrollo acelerado de tecnologías basadas en inteligencia artificial y, por otro, los valores comunicacionales, éticos y axiológicos históricamente asociados a la voz como expresión humana. La clonación vocal y los deepfakes de audio no constituyen fenómenos aislados, sino manifestaciones de un proceso más amplio de reconfiguración de la relación entre voz, tecnología y sociedad.

Desde una perspectiva comunicacional, la encrucijada digital de la voz se expresa en la progresiva pérdida de su función como garante de autenticidad y presencia. Tradicionalmente, la voz ha operado como un signo confiable de identidad y autoría, permitiendo a los interlocutores establecer relaciones basadas en la confianza y el reconocimiento mutuo. Sin embargo, la posibilidad de reproducir voces humanas de manera artificial introduce un escenario de ambigüedad en el que resulta cada vez más difícil distinguir entre lo humano y lo sintético. Esta situación afecta directamente los procesos de interpretación, credibilidad y sentido que sostienen la comunicación interpersonal y pública.

En términos axiológicos, la encrucijada digital pone de manifiesto una tensión entre la innovación tecnológica y el respeto por la dignidad humana. La conversión de la voz en un objeto técnico replicable tiende a despojarla de su valor simbólico y relacional, subordinándola a lógicas de eficiencia, control y rendimiento. Desde la óptica heideggeriana, esta transformación puede interpretarse como una

expresión del Gestell, en tanto la voz es reducida a un recurso disponible (Bestand), susceptible de ser almacenado, reproducido y explotado sin consideración por su arraigo existencial.

Asimismo, la encrucijada digital de la voz revela una crisis ética vinculada al consentimiento y la responsabilidad comunicativa. La utilización de voces clonadas o manipuladas sin autorización cuestiona principios fundamentales como la autonomía del sujeto y el derecho a la identidad. En este contexto, la voz deja de ser una extensión legítima del hablante para convertirse en un artefacto que puede circular independientemente de su voluntad, generando daños simbólicos, emocionales y sociales. Esta situación exige una reflexión crítica sobre los límites del uso tecnológico de la voz y sobre la necesidad de marcos normativos y éticos que protejan su valor humano.

Por otra parte, desde el campo de las Ciencias de la Comunicación, la noción de encrucijada digital permite integrar las dimensiones técnica, social y filosófica del problema. La voz se presenta simultáneamente como objeto de innovación, como recurso comunicacional vulnerable y como expresión fundamental de la subjetividad. Esta convergencia obliga a superar enfoques reduccionistas y a adoptar una mirada interdisciplinaria capaz de comprender la complejidad del fenómeno. En este sentido, la encrucijada digital no representa únicamente un riesgo, sino también una oportunidad para repensar el lugar de la voz en la comunicación contemporánea.

Ahora bien, la encrucijada digital de la voz evidencia un momento crítico en el que se decide si la tecnología contribuirá a enriquecer la experiencia comunicacional humana o si, por el contrario, profundizará procesos de deshumanización y pérdida de sentido. Reconocer esta encrucijada constituye un paso indispensable para avanzar hacia una gestión ética de las tecnologías vocales, orientada a preservar la voz como un bien comunicacional, relacional y axiológico en la era digital.

El reconocimiento automático del habla

El reconocimiento automático del habla (Automatic Speech Recognition, ASR) constituye una de las tecnologías fundamentales en el ecosistema contemporáneo de la inteligencia artificial aplicada a la comunicación. Estos sistemas están diseñados para convertir la señal vocal humana en texto procesable por máquinas mediante el análisis de las propiedades acústicas del habla. En términos generales, el reconocimiento de voz consiste en transformar las ondas acústicas del habla en secuencias de unidades lingüísticas, proceso que permite convertir el discurso oral en texto interpretable por sistemas informáticos (Jurafsky y Martin, 2023, p. 354). Esta tecnología se ha integrado de forma creciente en asistentes virtuales, sistemas de atención al cliente, dispositivos móviles y plataformas digitales.

De acuerdo a una perspectiva comunicacional, el reconocimiento automático del habla introduce una transformación significativa en la relación entre voz y escucha. Tradicionalmente, la escucha ha sido entendida como una práctica humana situada, mediada por la empatía, la contextualización y la interpretación subjetiva. Con el ASR, la escucha se automatiza y se desplaza hacia sistemas técnicos que procesan la voz como un conjunto de datos cuantificables. Este desplazamiento resulta central para el análisis desarrollado en este artículo, ya que implica que la voz no solo es intervenida en su producción —como ocurre con la clonación vocal—, sino también en su recepción e interpretación.

Por otro lado, en términos axiológicos, el ASR contribuye a la progresiva datificación de la voz humana. Al transformar el habla en información almacenada y analizable, estos sistemas facilitan prácticas de monitoreo, clasificación y evaluación que exceden la intención comunicativa original del hablante. En este sentido, Shoshana Zuboff (2019) sostiene que la captura sistemática de datos personales se inscribe en una lógica de explotación informacional que redefine la

experiencia humana como materia prima para procesos de predicción y control. Si bien su análisis no se centra específicamente en los datos vocales, documenta cómo diversas tecnologías digitales pueden incluso activar micrófonos y registrar el habla de los usuarios, lo que permite extender su planteamiento al ámbito de la voz . Desde esta perspectiva, su enfoque resulta útil para comprender los riesgos éticos asociados a la automatización de la escucha.

Con respecto al marco filosófico desarrollado previamente, el reconocimiento automático del habla puede interpretarse como una manifestación del Gestell heideggeriano, en tanto reduce la voz a un recurso disponible (Bestand) susceptible de ser procesado, almacenado y reutilizado. En este contexto, la voz deja de ser comprendida como acontecimiento relacional y expresivo del Dasein para convertirse en un insumo técnico dentro de sistemas orientados a la eficiencia y la optimización. Esta reducción técnica profundiza la separación entre voz, cuerpo e identidad, ya evidenciada en la clonación vocal y los deepfakes de audio.

Asimismo, el ASR desempeña un papel estructural en el desarrollo de tecnologías de síntesis y manipulación vocal. La capacidad de segmentar, analizar y modelar el habla humana constituye la base sobre la cual se construyen sistemas capaces de replicar voces con alta fidelidad. Es por ello, que el reconocimiento automático del habla no puede ser considerado un fenómeno aislado, sino un componente esencial del entramado tecnológico que interviene la voz en la actualidad. Esta interdependencia refuerza la necesidad de analizar el ASR desde una perspectiva ética y comunicacional, y no únicamente desde criterios de rendimiento técnico.

En ese sentido, el reconocimiento automático del habla amplía y profundiza la encrucijada digital de la voz al automatizar no solo su producción, sino también su escucha e interpretación. La transformación de la voz en datos procesables plantea desafíos significativos para la ética de la comunicación, al cuestionar los

límites entre comprensión humana, control tecnológico y pérdida de la dimensión relacional del acto comunicativo. Integrar el análisis del ASR permite, por tanto, comprender de manera más completa cómo la tecnificación de la voz afecta los valores fundamentales que sostienen la comunicación humana en la era digital.

Consideraciones metodológicas

El presente artículo se inscribe dentro de un enfoque cualitativo de investigación, orientado a la comprensión e interpretación de fenómenos comunicacionales contemporáneos vinculados a la voz humana en contextos tecnológicos avanzados. Dada la naturaleza teórica y reflexiva del objeto de estudio, se adoptó como estrategia metodológica la revisión documental, entendida como un proceso sistemático de análisis de fuentes académicas relevantes para la construcción del marco conceptual y argumentativo del trabajo.

Por otra parte, la revisión documental se realizó a partir de libros especializados, artículos científicos y publicaciones académicas recientes, principalmente del período comprendido entre 2020 y 2025, provenientes de los campos de las Ciencias de la Comunicación, la ética, la filosofía de la tecnología, la inteligencia artificial y las ciencias de la voz. Esta selección permitió identificar categorías clave como voz humana, clonación vocal, deepfakes de audio, reconocimiento automático del habla, transhumanismo y axiología comunicacional, las cuales estructuran el desarrollo teórico del artículo.

Desde el punto de vista epistemológico, el análisis de las fuentes se abordó mediante un enfoque hermenéutico–interpretativo, en la línea de la fenomenología hermenéutica de Heidegger, centrada en la comprensión como modo fundamental de ser-en-el-mundo, en tanto se privilegió la comprensión crítica de los discursos teóricos y científicos en su contexto de producción. La hermenéutica permitió interpretar los textos no como unidades aisladas, sino como expresiones de marcos conceptuales

más amplios que dialogan entre sí y reflejan tensiones contemporáneas entre tecnología, comunicación y valores humanos. Este enfoque resulta particularmente pertinente para el estudio de la voz, entendida como un fenómeno que articula dimensiones biológicas, simbólicas y sociales.

Ahora bien, el procedimiento metodológico consistió en la lectura analítica, comparación y articulación de las fuentes seleccionadas, con el fin de identificar convergencias, divergencias y vacíos teóricos en torno a la intervención tecnológica sobre la voz humana. A partir de este proceso, se construyó un recorrido conceptual progresivo, que parte de la voz como fenómeno humano y comunicacional, para luego analizar su transformación en el contexto de la inteligencia artificial, la clonación vocal, los deepfakes de audio y el reconocimiento automático del habla, culminando en una reflexión ética y axiológica situada en el marco del transhumanismo.

En este sentido, el artículo no busca establecer generalizaciones empíricas ni mediciones cuantitativas, sino problematizar y comprender críticamente un fenómeno emergente que impacta de manera directa los fundamentos de la comunicación humana. La elección de la revisión documental y del enfoque hermenéutico responde, por tanto, a la necesidad de abordar la complejidad del objeto de estudio desde una perspectiva interpretativa, coherente con los objetivos y con el campo de las Ciencias de la Comunicación.

Reflexiones finales

El análisis desarrollado a lo largo de este artículo permite afirmar que la voz humana ocupa un lugar central en la configuración de los procesos comunicacionales, no solo como medio de transmisión de información, sino como expresión integral de la identidad, la presencia y la relación con el otro. Comprendida desde una perspectiva sistémica —fisiológica, comunicacional y axiológica—, la voz se constituye en un bien comunicacional de alto valor humano, cuya autenticidad ha sustentado históricamente la confianza social y la

legitimidad de los intercambios discursivos.

Sin embargo, el desarrollo acelerado de tecnologías basadas en inteligencia artificial ha introducido una transformación profunda en esta relación. La clonación de voz, los deepfakes de audio y el reconocimiento automático del habla evidencian un proceso de intervención tecnológica que disocia la voz del cuerpo y del sujeto que la produce, convirtiéndola en un objeto técnico replicable, analizable y manipulable. Esta disociación no solo altera las formas de producción y circulación de mensajes, sino que impacta directamente los valores que sostienen la comunicación humana, dando lugar a una crisis de credibilidad sonora que atraviesa tanto las interacciones interpersonales como los discursos públicos e institucionales.

Desde una perspectiva filosófica, el transhumanismo ofrece un marco interpretativo relevante para comprender estas transformaciones, al proponer la optimización y sustitución de capacidades humanas mediante la tecnología. No obstante, el análisis realizado pone en evidencia que, aplicada al ámbito comunicacional, esta lógica entra en tensión con la dimensión ética y axiológica de la voz. La búsqueda de eficiencia y escalabilidad tecnológica corre el riesgo de reducir la voz a un conjunto de datos, despojándola de su carácter relacional y de su función como garante simbólico de autenticidad y responsabilidad comunicativa.

En este contexto, la encrucijada digital en la que se sitúa la voz humana no puede ser abordada únicamente desde criterios técnicos o normativos. Se requiere una reflexión ética profunda que reconozca la voz como una expresión irreductible de la dignidad y la identidad del sujeto. La automatización de la producción y la escucha vocal plantea desafíos que obligan a repensar los límites del desarrollo tecnológico, así como la necesidad de establecer principios orientados a una gestión responsable de la innovación, centrada en el respeto por lo humano.

Desde el campo de las Ciencias de la Comunicación, este trabajo propone comprender la problemática de la voz

en la era digital como un fenómeno complejo que exige un enfoque interdisciplinario, capaz de articular saberes provenientes de la vocología, la filosofía, la ética y la comunicación. La revisión documental y el análisis hermenéutico realizados permiten concluir que la preservación del valor axiológico de la voz constituye un desafío ineludible frente a las tendencias poshumanistas emergentes.

Finalmente, este artículo invita a abrir nuevas líneas de investigación orientadas a profundizar en los impactos sociales, culturales y éticos de las tecnologías vocales, así como en el diseño de marcos regulatorios y educativos que promuevan una relación más equilibrada entre tecnología y humanidad. Garantizar que el desarrollo técnico se articule con los valores fundamentales de la comunicación humana no es solo un desafío académico, sino una responsabilidad ética frente al futuro de la vida social en entornos digitales cada vez más complejos.

REFERENCIAS BIBLIOGRÁFICAS

Bostrom, N. (2005). Historia del pensamiento transhumanista. Journal of Evolution and Technology, 14(1), 1–25. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://jetpress.org/volume14/bostrom.pdf

Chesney, R., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. California Law Review, 107(6), 1753–1820. https://doi.org/10.15779/Z38RV0D15J

Couldry, N. (2010). Why voice matters: Culture and politics after neoliberalism. Sage.

Couldry, N. (2012). Media, society, world: Social theory and digital media practice. Polity Press.

Couldry, N., & Hepp, A. (2017). The mediated construction of reality. Polity Press.

de Ruiter, A. (2021). The distinct wrong of deepfakes. Philos. Technol. 34 , 1311–1332. https://doi.org/10.1007/s13347-021-00459-2

Europol. (2023). Facing reality? Law enforcement and the challenge of deepfakes. Publications Office of the European Union.

https://www.europol.europa.eu

Fairclough, N. (2010). Critical discourse analysis: The critical study of language (2nd ed.). Routledge. https://www.felsemiotica.com/descargas/Fairclough-Norman-Critical-Discourse-Analysis.-The-Critical-Study-of-Language.pdf

Franganillo, J. (2023). La inteligencia artificial generativa y su impacto en la creación de contenidos mediáticos. methaodos.revista de ciencias sociales, 11(2), m231102a10. http://dx.doi.org/10.17502/mrcs.v11i2.710

Heidegger, M. (1994). La pregunta por la técnica (E. Barjau, Trad.). Alianza. (Trabajo original publicado en 1954).

Heidegger, M. (2006). Ser y tiempo. Disponible en: https://apiperiodico.jalisco.gob.mx/api/sites/periodicooficial.jalisco.gob.mx/files/ser_y_tiempo-martin_heidegger.pdf.

Huxley, J. (1957). New bottles for new wine. Chatto & Windus. https://ia801705.us.archive.org/2/items/NewBottlesForNewWine/New-Bottles-For-New-Wine.pdf

Jurafsky, D., & Martin, J. H. (2023). Speech and Language Processing (3rd ed. draft). Stanford University. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://web.stanford.edu/~jurafsky/slp3/ed3bookaug20_2024.pdf

Luhmann, N. (2005). Confianza. Anthropos Editorial.

Poyatos, F. (2002). Nonverbal communication across disciplines. Vol. II: Paralanguage, kinesics, silence, personal and environmental interaction. John Benjamins.

Sukumar, M., Kodanda, P., Raga, S., & Lalitha, S. (2024). Speech and speaker recognition approaches: A survey. International Journal of Electrical and Computer Engineering, 14(1), 338–350. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.ijcrt.org/papers/IJCRT2410620.pdf

Titze, I. R. (2000). Principles of voice production. National Center for Voice and Speech.

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://journals.sagepub.com/doi/epub/10.1177/2056305120903408

Van Leeuwen, T. (1999). Speech, Music, Sound. Houndmills y Londres: MacMillan. 231 págs.

Zhang B, Cui H, Nguyen V, Whitty M. (2025). Audio Deepfake Detection: What Has Been Achieved and What Lies Ahead. Sensors (Basel). 22;25(7):1989. doi: 10.3390/s25071989. PMID: 40218502; PMCID: PMC11991371. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://pmc.ncbi.nlm.nih.gov/articles/PMC11991371/pdf/sensors-25-01989.pdf

Zuboff, S. (2019). The age of surveillance capitalism. PublicAffairs.