

# Voces sin rostro. Los retos de la inteligencia artificial en la industria del doblaje

## *Faceless Voices. The Challenges of Artificial Intelligence in the Dubbing Industry*

Hugo Ignacio Buendía Romero

 <https://orcid.org/0009-0000-2741-2206>

Universidad Nacional Autónoma de México, México

Correo electrónico: hbrlegalstudio@gmail.com

Recepción: 12 de junio de 2025

Aceptación: 9 de diciembre de 2025

Publicación: 29 de enero de 2026

DOI: <https://doi.org/10.22201/ijj.24484873e.2025.174.20248>

**Resumen:** El artículo examina cómo las tecnologías contemporáneas de síntesis de voz tensionan las categorías jurídicas tradicionales aplicables a la voz humana y el doblaje, particularmente en el régimen de derechos de autor, derechos conexos y derechos de la personalidad. A través de una aproximación jurídico-dogmática apoyada en un análisis técnico accesible, se distinguen cuatro familias tecnológicas de generación y transformación vocal y se estudian casos relevantes que permiten identificar riesgos y vacíos regulatorios emergentes. Los resultados de esta revisión muestran tres retos jurídicos centrales: la fragilidad del consentimiento frente a la extracción y reutilización algorítmica de la voz; el desplazamiento laboral derivado de la sustitución técnica de equipos y actores de doblaje, y la incertidumbre sobre la titularidad y naturaleza jurídica del clon vocal producido por sistemas de inteligencia artificial. Asimismo, se plantean posibles principios que podrían guiar respuestas regulatorias a dichos retos.

**Palabras clave:** doblaje; *AI dubbing*; clonación de voz; derechos conexos; derechos de imagen; voz como dato biométrico.

**Abstract:** The article examines how contemporary voice synthesis technologies challenge traditional legal categories applicable to the human voice, interpretation, and dubbing, particularly in the areas of copyright, related rights, and personality rights. Through a legal-dogmatic approach supported by accessible technical analysis, four technological families of voice generation and transformation are distinguished, and relevant cases are studied to identify emerging risks and regulatory gaps. The results of this review reveal three central legal challenges: the fragility of consent in the face of algorithmic voice extraction and reuse; job dis-

placement resulting from the technical replacement of interpreters and dubbing actors; and uncertainty about the ownership and legal nature of voice clones produced by artificial intelligence systems. Possible principles that could guide regulatory responses to these challenges are also proposed.

**Keywords:** dubbing; AI dubbing; voice cloning; neighbouring rights; image rights; voice-as-biometric-data.

**Sumario:** I. *Introducción*. II. *Industria del doblaje*. III. *Naturaleza jurídica del doblaje*. IV. *La voz como derecho*. V. *Voces artificiales*. VI. *AI dubbing*. VII. *Casos de estudio*. VIII. *Retos y el futuro del doblaje ante el AI dubbing*. IX. *Conclusiones*. X. *Referencias*.

## I. Introducción

Hasta hace pocos años, el doblaje implicaba un proceso artesanal en donde un actor humano grababa su voz en un estudio, el ingeniero de sonido editaba y sincronizaba la pista de audio, y el producto final reemplazaba la pista original para quienes consumían contenido en otro idioma. Sin embargo, con la aparición de tecnologías basadas en redes neuronales —TACOTRON 2, StarGAN-VC, AutoVC, FlowDubber, etcétera— es posible generar réplicas hiperrealistas de voces humanas en cuestión de segundos. Estas “voces sintéticas” pueden modularse, clonarse y distribuirse sin la mediación directa de un actor. La irrupción del *AI dubbing* en la industria del doblaje no sólo plantea desafíos técnicos, sino que pone en crisis algunas de las categorías jurídicas más consolidadas del derecho de autor. ¿Qué ocurre cuando la voz humana, caracterizada por ser intrínsecamente personal y expresiva, se convierte en insumo para ser clonada? ¿Qué ocurre cuando una máquina no sólo traduce, sino que interpreta? ¿Cómo se redefine el uso de una obra o incluso de una voz cuando se abstrae y transforma algorítmicamente?

Este artículo se inscribe en la encrucijada de ese debate jurídico. Por un lado, la voz humana ha sido considerada un atributo de la personalidad y un objeto de protección, tanto en el derecho de autor como en los derechos conexos cuando se plasma en una obra. Por otro, el advenimiento de la IA generativa introduce un cambio de paradigma: la voz deja de “fijarse” en un soporte humano para convertirse en un insumo algorítmico. La pre-

sente investigación analiza la interacción entre las tecnologías de síntesis y transformación de la voz (TTS generativo, *voice conversion*, *voice cloning* y los *AI dubbing pipelines*) y el régimen jurídico del derecho de autor, y revisa su desarrollo técnico y los problemas que plantean para sus categorías. El estudio incorpora la dimensión de la voz como atributo de la personalidad y como dato biométrico, al atender a los riesgos derivados tanto de su utilización como material de entrenamiento, como de la generación algorítmica de salidas que reproducen estilos o identidades vocales de personas determinadas. La hipótesis central es que la síntesis algorítmica vocal desancla a la voz de su corporeidad y coloca al intérprete en un régimen de vulnerabilidad jurídica y laboral no atendido por el derecho vigente. Asimismo, se proponen principios que podrían orientar respuestas regulatorias frente a estos retos.

### Metodología y objetivos

La metodología de este trabajo articula un enfoque analítico-descriptivo con un componente técnico-exploratorio y una revisión contextual de precedentes internacionales. El enfoque jurídico-dogmático permite interpretar doctrina, principios y regímenes aplicables a la voz, y el doblaje, así como evaluar su capacidad para responder a fenómenos tecnológicos emergentes. El componente técnico-descriptivo se apoya en literatura especializada en síntesis y transformación de voz para delimitar las cuatro familias tecnológicas estudiadas (síntesis tradicional, clonación vocal, conversión de voz y sistemas inteligentes de doblaje) y precisar sus efectos relevantes para la protección jurídica de la interpretación y la identidad vocal. Finalmente, la revisión contextual de precedentes internacionales se utiliza como método de observación indirecta, no para establecer comparaciones normativas estrictas, sino para identificar patrones, tensiones y riesgos comunes en litigios o controversias derivados del uso de estas tecnologías.

A partir de esta metodología, el análisis se organiza en tres niveles: *a*) caracterización técnica y jurídica del doblaje humano como punto de partida; *b*) delimitación conceptual y funcional de las tecnologías de síntesis y transformación vocal, y *c*) tensiones teóricas y estudio de casos representativos (Roadrunner, SAG-AFTRA vs. Epic Games y Yin vs. Beijing Internet Court) que permiten observar cómo dichas tecnologías operan en contextos reales

y qué desafíos plantean para el derecho de autor, los derechos conexos y los derechos de la personalidad. Esta articulación metodológica hace posible formular los problemas centrales del fenómeno y proponer principios que puedan ser incorporados en la solución a dichos retos.

## II. Industria del doblaje

El doblaje constituye una de las formas más consolidadas de traducción audiovisual y se ha extendido ampliamente en el ámbito internacional, con especial arraigo en países como Alemania, España, Japón, Brasil o Francia. Sus orígenes se sitúan a finales de los años veinte, cuando la aparición de las películas sonoras (llamados *talkies*) exigió nuevas formas de traducción, como la sustitución de intertítulos por doblaje o subtítulo en lenguas como el francés, alemán y español (Bosseaux, 2018). Ávila (1997, p. 18) lo define como “la grabación de una voz en sincronía con los labios de un actor de imagen o una referencia determinada, que imite lo más fielmente posible la interpretación de la voz original”. Su objetivo es producir la impresión de que los personajes en pantalla hablan naturalmente en el idioma de destino (Luyken, 1991), a fin de que esta mediación vocal permita traducir y hacer inteligibles imágenes, sonidos, gestos, emociones e ideas entre individuos y entre naciones (Nájar, 2007). Se trata, entonces, de un proceso intrínsecamente complejo que integra dimensiones lingüísticas y culturales; pero también requerimientos técnicos y artísticos (Chaume, 2012).

Vale la pena señalar que el doblaje es sustancialmente distinto del subtítulo, el cual consiste en incorporar un texto escrito (subtítulos) en el idioma de destino en la pantalla donde se muestra la versión original (Ivarsson, 1998); o del *voice over*, práctica en donde la pista de audio con la grabación del diálogo original se reproduce a un menor volumen, al mismo tiempo que la pista con la versión traducida (Chaume, 2012). El doblaje pues, supone una sustitución total del canal de voz original (Indira Gandhi National Open University, s. f.).

Zabalbeascoa (1997) destaca que elementos como el lenguaje corporal, las expresiones faciales, los titubeos y el contacto visual influyen decisivamente en cómo un actor de doblaje capta e interpreta el sentido que un hablante

pretende transmitir. Estos mismos factores no verbales intervienen directamente en la traducción audiovisual, y condicionan la forma en que el mensaje debe ser trasladado a otra lengua. En ese sentido, y de forma general, Toda señala que “los condicionamientos que tiene la traducción audiovisual la diferencian de otras modalidades, convirtiéndola en una variedad especializada, distinta de la traducción literaria” (2005, p. 119). En consecuencia, el resultado de este proceso no es fruto del azar ni de una ejecución automática, sino de una compleja cadena de producción simbólica, compuesta de múltiples etapas íntimamente relacionadas, y que involucra a diversos participantes como traductores, ajustadores, actores de voz, directores y técnicos especializados (Martínez, 2004). Cada uno de ellos contribuye a que el producto final pueda ser aceptado por una audiencia como una ilusión creíble (Matamala, 2010). El doblaje, en su mejor expresión, es una operación de orfebrería simbólica; un minucioso trabajo en el que cada decisión puede alterar por completo el éxito o fracaso de la recepción de una obra. Esta sofisticación no se limita al ámbito cinematográfico. Como señala Nájjar (2007), el doblaje también se emplea en publicidad, en videojuegos y materiales digitales que requieren sonorización sincronizada, en dramatizaciones radiofónicas e, incluso, en espectáculos en vivo.

Resulta pertinente entonces, esbozar una breve aproximación a los sujetos y procedimientos intervinientes en este proceso, con el propósito de vislumbrar con mayor nitidez la complejidad que entraña la producción del doblaje. Martínez (2004) señala que el proceso de doblaje se activa en la fase de preproducción, cuando el cliente entrega al estudio la copia maestra, el guión y las instrucciones técnicas. El jefe de producción envía este material al traductor, quien puede lidiar con guiones incompletos o desalineados con la obra audiovisual. Una vez traducido, el texto suele pasar por una revisión lingüística especializada. Posteriormente, apunta Matamala (2010), el texto se sincroniza con las imágenes por otro profesional o por el mismo traductor. En esta fase se ajusta a los movimientos labiales y al ritmo visual mientras se procura conservar el sentido original. Finalmente, el texto sincronizado pasa al departamento de producción, donde se prepara formalmente para su grabación en cabina. Zabalbeascoa (1997) resalta por su parte que, detalles como

el momento en que se pronuncia (*timing*) es un factor importante en el doblaje, tanto para el movimiento de los labios como para ciertas situaciones que requieran respuestas especiales con una sincronización especial, por ejemplo, un chiste o un momento dramático culminante. (p. 336)

Una vez iniciada la sesión de grabación, intervienen diversos participantes que asumen funciones altamente técnicas y decisivas para lograr una interpretación efectiva de la obra. Tal como expone Luyken (1991), este momento implica la proyección del programa, la entrega del guión traducido, la actuación de los intérpretes bajo la guía del director y la grabación final del *performance*. Al retomar la clasificación propuesta por Hernández Reséndiz (1996), estos roles pueden agruparse del siguiente modo: el actor de doblaje, profesional con formación actoral cuya precisión interpretativa, dicción y capacidad de sincronía sostienen la creación vocal; el director de doblaje, responsable del *casting*, de guiar la entonación y de asegurar la correspondencia entre la voz grabada y la imagen original; el operador de sonido, técnico que graba, edita e integra la pista doblada al material audiovisual para garantizar una sincronización exacta; y el productor, quien coordina el proyecto en su conjunto, gestiona los recursos técnicos y artísticos y contrata al equipo necesario para cumplir los estándares de calidad y los plazos de entrega.

En consecuencia de todo lo esgrimido, y como señala Palencia, “el doblaje elimina por completo un elemento de la obra original, que son las voces de los actores que los consumidores de productos subtítulos, en cambio, sí perciben” (2002, p. 20). Luyken (1991) observa que, en determinadas cinematografías, la voz del actor de doblaje puede llegar a asociarse de manera tan estrecha con la figura de una estrella que el público rechaza cualquier sustitución, incluso cuando se trata de producciones distintas. Este fenómeno ilustra hasta qué punto el doblaje ha evolucionado hasta convertirse en un componente simbólico, cultural y económico clave en la circulación internacional de contenidos audiovisuales. Esto da paso a revisar la naturaleza jurídica de los productos provenientes de las labores especializadas implicadas en el doblaje; pues, en efecto, más allá del mero traslado lingüístico, el doblaje constituye un proceso interpretativo y técnico complejo que genera resultados autónomos respecto de la obra primigenia. La voz del actor de doblaje no es un mero canal neutro; imprime matices, emociones y modulaciones que inci-

den directamente en la recepción estética y narrativa de la obra audiovisual traducida. Por ello, dentro del régimen de derecho de autor, estas contribuciones no se califican como obras nuevas, sino como interpretaciones protegidas por derechos conexos.

### III. Naturaleza jurídica del doblaje

Mientras que los derechos de autor protegen a los creadores de obras, los derechos conexos protegen a determinados intermediarios, es decir, a algunos de los responsables de que las obras lleguen al público. Magaña (2019), sugiere que estos derechos son una retribución para los artistas, intérpretes, ejecutantes, editores, productores y transmisores de una obra autoral por su participación, actuación, interpretación, promoción o difusión de la obra autoral. En este marco, el objeto protegido es la actuación personal del artista intérprete o ejecutante en un soporte material (Lipszyc, 2017), pues la voz, como fenómeno sonoro y atributo de la personalidad, no es protegible *per se* por el derecho de autor, sino únicamente en la medida en que forma parte de una obra efectivamente fijada.

Al igual que en materia de derechos de autor, el contenido de los derechos conexos de los artistas son las distintas facultades, tanto de tipo económico como de tipo moral, que se generan automáticamente con la fijación de la actuación y se pueden ejercer sobre las actuaciones de los propios artistas sin perjuicio de la titularidad de la obra audiovisual original (Parra Trujillo, 2015). Al respecto, Rangel Medina precisa que, “el intérprete debe frenar su personalidad creadora para subordinarla a la del autor, porque si da autonomía a su interpretación, deja de ser esta para convertirse en una adaptación, que requiere una autorización especial del autor” (1998, p. 124). Bajo esa misma línea argumentativa, la Convención de Roma sobre la protección de los artistas intérpretes o ejecutantes, los productores de fonogramas y los organismos de radiodifusión (1961) señala que, se entenderá por: “a) «artista intérprete o ejecutante», todo actor, cantante, músico, bailarín u otra persona que represente un papel, cante, recite, declame, interprete o ejecute en cualquier forma una obra literaria o artística [...]”.

Se deduce de la anterior transcripción que resulta lógico encuadrar que la labor realizada por los actores de doblaje constituye la de un artista intérprete o ejecutante, puesto que, mediante el uso de su voz, representan personajes fijados en una obra audiovisual e interpretan guiones o *screenplays* previamente incluidos en la misma y transmiten así, con intención dramática, el contenido expresivo del texto original. En este sentido, Lipszyc subraya que, “los intérpretes realizan actividades personales de carácter artístico, motivo por el cual las legislaciones establecen a su favor facultades concernientes a la tutela de su personalidad” (2017, p. 243). Así, la Organización Mundial de la Propiedad Intelectual (2016) concuerda en que dicho reconocimiento se justifica en la medida en que se considera necesaria su intervención creativa.

De manera general, el derecho moral de los artistas intérpretes o ejecutantes comprende dos facultades principales: la de paternidad, que les garantiza el reconocimiento de su nombre en relación con sus interpretaciones; y la de integridad, que les permite oponerse a cualquier alteración de su actuación que pueda afectar su prestigio o reputación (Parra Trujillo, 2015). Asimismo, en el plano económico, esta condición, según lo establecido por la Convención de Roma, otorga a los artistas intérpretes una serie de derechos:

La protección prevista por la presente Convención en favor de los artistas intérpretes o ejecutantes comprenderá la facultad de impedir:

- a) la radiodifusión y la comunicación al público de sus interpretaciones o ejecuciones para las que no hubieren dado su consentimiento, excepto cuando la interpretación o ejecución utilizada en la radiodifusión o comunicación al público constituya por sí misma una ejecución radiodifundida o se haga a partir de una fijación;
- b) la fijación sobre una base material, sin su consentimiento, de su ejecución no fijada;
- c) la reproducción, sin su consentimiento, de la fijación de su ejecución:
  - (i) si la fijación original se hizo sin su consentimiento;
  - (ii) si se trata de una reproducción para fines distintos de los que habían autorizado;
  - (iii) si se trata de una fijación original hecha con arreglo a lo dispuesto en el artículo 15 que se hubiera reproducido para fines distintos de los previstos en ese artículo.

De ahí que, tanto el proceso de doblaje como los resultados interpretativos que genera (las fijaciones sonoras y audiovisuales que encarnan la aportación creativa del intérprete), queden amparados por el sistema dual de protección del derecho de autor y los derechos conexos. El primero, en cuanto tutela la obra originaria adaptada; y los segundos, en cuanto reconocen y procuran la protección a la dimensión creativa, personalísima y diferenciable de la interpretación vocal fijada.

Ahora bien, comprender este armazón jurídico exige no perder de vista las condiciones materiales en que dichas interpretaciones se producen, pues la eficacia real de la protección normativa se ve erosionada cuando los intérpretes operan en contextos marcados por la precariedad. En consecuencia, resulta prudente realizar una breve aproximación hacia el panorama laboral dentro de la industria del doblaje, pues es un hecho que sectores como el actuarial enfrentan altos niveles de precariedad e inestabilidad, incluso en países con industrias culturales consolidadas. En España, el Informe Sociolaboral 2023 de AISGE (Fundación AISGE, 2024) muestra que 48% de los profesionales obtiene menos de 3,000 euros anuales por su labor actuarial, lo que obliga a más de la mitad a buscar empleos adicionales y los mantiene en condiciones cercanas al umbral de pobreza. En el Reino Unido, el estudio de la University of Exeter (Pavis, Tulti y Pye, 2019) identifica un mercado caracterizado por baja formalización contractual, escaso conocimiento de derechos de PI, débil representación sindical, fragilidades exacerbadas por plataformas *peer-to-peer* configurando un escenario particularmente permeable a los desarrollos tecnológicos recientes, incluida la incorporación de sistemas de inteligencia artificial. El panorama argentino refleja patrones similares: remuneraciones promedio de apenas 1,500 pesos por episodio (cerca de 7.5 dólares como único pago), creciente informalidad y presión por abaratar costos, condiciones que, según Fleischer (2022), han normalizado prácticas laborales que comprometen derechos básicos de actores y actrices de voz.

Esta situación no sólo exhibe la fragilidad estructural de la seguridad económica del sector, sino que también revela una exposición creciente a transformaciones tecnológicas capaces de amplificar la ya volátil arquitectura de incertidumbres que atraviesa.

## IV. La voz como derecho

En un plano más elemental, la voz humana nace de un proceso estrictamente corporal. La energía cinética del aire que asciende desde los pulmones impulsa la vibración rítmica de las cuerdas vocales, cuya apertura y cierre alternado produce el flujo glotal necesario para que emerjan los sonidos que conforman el timbre singular de cada individuo (Chen, 2016). Cavarero afirma incluso que, “desde la escena materna en adelante, la voz manifiesta el ser único de cada ser humano y su autocomunicación espontánea según los ritmos de una relación sonora” (2005, p. 173). Esta dinámica, determinada materialmente por la fisiología de cada individuo, resulta entonces indisoluble de la identidad corporal y expresiva de la persona. Hecho que, por consiguiente, sugiere que toda voz es, antes que un signo comunicativo, una emanación material y única cargada de matices emocionales, sociales y culturales. En este contexto, se vuelve necesario examinar cómo el derecho ha abordado la voz no sólo como medio de expresión artística, sino también como atributo de la personalidad protegido en el marco del derecho de imagen y del dato biométrico.

A partir de estas coordinadas, doctrinarios como Flores Ávalos y Pérez García definen el derecho de imagen “como el derecho de la personalidad —por ende, derecho subjetivo— que faculta a la persona para impedir que su apariencia física y/o su voz sean reproducidas de cualquier manera si ella no otorga su consentimiento” (2019, p. 5). Por su parte, Nogueira Alcalá (2007, p. 261) define que, esto sucede “en la medida que expresan cualidades morales de la persona y emanaciones concretas de su dignidad de ser humano, configurando su ámbito personal e instrumento básico de su identificación, proyección exterior y reconocimiento como ser humano”. En este sentido, el derecho a la imagen se extiende a toda forma de representación, captación, reproducción o difusión de la figura o voz humanas, sin importar el medio utilizado, su propósito o duración en el tiempo. Esta facultad confiere a su titular un control exclusivo sobre el uso de dicha representación, siempre dentro de los márgenes establecidos por la ley (Trujillo Cabrera, 2024).

Paralelamente, la doctrina ha subrayado que la voz opera también como un atributo biométrico. Jasserand (2016) explica que características biométricas como el rostro, las huellas digitales o la voz comparten rasgos distintivos

y repetibles que permiten identificar a un individuo mediante la medición y comparación de patrones específicos. En una línea similar, Pizzolo (2025) sostiene que los datos biométricos transforman radicalmente la relación entre cuerpo e identidad al volver mecánicamente legibles las cualidades fisiológicas y conductuales de la persona, permitiendo así su identificación unívoca, pero exponiéndola al mismo tiempo a riesgos agravados cuando dicha información se utiliza de forma indebida.

Con ello, se amplía el espectro de protección hacia componentes sonoros e intangibles que si bien no son visibles ni constituyen expresión artística por sí mismas, sí implican expresiones esenciales de la identidad personal. La voz, en particular, se configura como una huella acústica que permite reconocer y singularizar a un individuo. En consecuencia, al tratarse de una manifestación de la personalidad, resulta intransmisible, irrenunciable e inembargable. Por lo que, aunque se admita la posibilidad de explotar ciertas manifestaciones del individuo, lo que realmente ocurre es una cesión limitada de esos aspectos. No se trata de una disposición absoluta ni integral de los elementos que conforman el derecho a la propia imagen (Ordelín, 2023). Por ende, su apropiación indebida o uso no autorizado podría traducirse en conductas que lesionan la dignidad humana, al afectar directamente la esfera más íntima y representativa del sujeto.

Así pues, la centralidad de la voz como manifestación identitaria ha llevado a distintos ordenamientos a integrarle dentro del catálogo de atributos protegidos de la personalidad. Aunque los enfoques varían, emerge un consenso incipiente en torno a su carácter representativo y a la necesidad de evitar su apropiación o explotación no consentida:

- En Argentina, el Código Civil y Comercial en su artículo 53 exige claramente el consentimiento para captar o reproducir la imagen o la voz de una persona por cualquier medio.
- Chile reconoce expresamente la voz como dato personal biométrico sensible, al incluirla junto con el iris, las huellas digitales y los rasgos faciales, dentro de las características que permiten la identificación única de una persona (Ley 19.628, art. 16 *ter*).
- España contempla protección explícita a la voz como atributo propio de la persona mediante su inclusión en la Ley Orgánica 1/1982 de pro-

tección civil del derecho al honor, la intimidad personal y familiar y a la propia imagen.

- En Francia, el Tribunal de París resolvió que la protección de la voz permite a cualquier persona oponerse a su explotación comercial no autorizada, al constituir un atributo íntimamente ligado a la identidad y a la vida privada del individuo (Tribunal judiciaire de Paris, 2023).
- Aunque Estados Unidos no reconoce a nivel federal la voz como dato biométrico protegido, existen marcos estatales particularmente robustos como el Biometric Information Privacy Act de Illinois, que incluye expresamente la *voiceprint* dentro de los identificadores biométricos que imponen obligaciones estrictas de consentimiento, tratamiento y responsabilidad civil (2008).
- En México, la Suprema Corte de Justicia de la Nación en México ha señalado que el derecho a la propia imagen debe contemplar cualquier elemento representativo de la persona, sin limitarse estrictamente a sus rasgos físicos inmediatos. Esta interpretación amplia permite incorporar manifestaciones como la voz dentro del ámbito de protección de la personalidad (Tesis: 1a./J. 166/2023 (11a.), 2023).

De esta forma, ya sea bajo el derecho a la imagen, la privacidad, los datos biométricos o los derechos de la personalidad, emerge un consenso normativo que reconoce que la voz no es un mero sonido intercambiable, sino una manifestación singular anclada a la persona y a su dignidad. La sofisticación de herramientas de síntesis y transformación vocal introduce pues, tensiones inéditas entre innovación y tutela de la personalidad. De ello se colige que el uso indebido de la voz trasciende el plano patrimonial. Así, aunque este artículo no abordará de manera específica el fenómeno de los *deep fakes*<sup>1</sup> de voz, conviene tenerlo presente como un escenario paradigmático en el que estos riesgos se intensifican y se vuelven más visibles.

---

<sup>1</sup> “Systems enable the creation of hyper-realistic modifications to audio, video, or images by training neural networks to compete in producing increasingly convincing outputs” (Bhandari y Bhandari, 2025, p. 2).

## V. Voces artificiales

En este punto, resulta inevitable advertir que la comprensión jurídica de la voz como emanación corpórea, atributo de la personalidad (imagen) y vector indisociable de identidad, se tensiona de manera creciente frente al advenimiento de tecnologías capaces de desmaterializarla, modularla y replicarla con una precisión otrora impensable. Esta intersección marca un umbral en el que la voz deja de ser únicamente un fenómeno fisiológico para convertirse también en un objeto técnico manipulable.

La investigación en síntesis de voz tiene raíces que se remontan al siglo XVIII, cuando surgieron los primeros dispositivos conocidos como *têtes parlantes* o “cabezas parlantes”, intentos mecánicos rudimentarios por emular la voz humana (Parssinen, 2007). Desde entonces, la síntesis de voz, o *speech synthesis*, reúne un conjunto de tecnologías que reconfiguran la producción sonora mediante arquitecturas capaces de generar, imitar o transformar voces humanas. Desde los modelos TTS tradicionales hasta la clonación, la conversión y los sistemas integrados de *AI dubbing*, estas técnicas descomponen la voz en unidades acústicas manipulables y la reconstruyen como un simulacro verosímil de identidad vocal.

Antes de abordar la síntesis del habla como fenómeno que ha impactado la práctica del doblaje y consigo los derechos relacionados, conviene establecer algunas definiciones técnicas básicas que permitirán comprender el alcance de los procesos automatizados involucrados:

- *Machine learning*, o aprendizaje automático, consiste en programar computadoras hasta ciertos parámetros, el aprendizaje es la ejecución de un programa informático para optimizar los parámetros del modelo utilizando los datos de entrenamiento o la experiencia pasada (Alpaydm, 2014).
- *Deep learning*, o aprendizaje profundo es una subárea del aprendizaje automático que permite a los sistemas de inteligencia artificial extraer representaciones directamente a partir de datos, sin necesidad de que un humano defina manualmente las características relevantes, al aprender automáticamente jerarquías de representaciones a partir de datos crudos como imágenes, texto o audio. Esto le permite abordar tareas complejas donde las relaciones entre datos no son evidentes ni fácilmente

codificables, como el reconocimiento de patrones en imágenes médicas o el procesamiento del lenguaje natural (Goodfellow, Bengio, y Courville, 2016).

- GAN (*Generative Adversarial Networks*), las cuales son un tipo de modelo generativo basado en el aprendizaje profundo que consiste en dos redes enfrentadas: una generadora, que produce datos sintéticos a partir de ruido, y una discriminadora, que distingue entre datos reales y generados. Ambas se entrenan de forma simultánea en un esquema competitivo que permite a la generadora aproximar distribuciones complejas y crear salidas cada vez más verosímiles, sin requerir cálculos probabilísticos exhaustivos (Goodfellow *et al.*, 2014).
- *Zero-shot learning*, método de entrenamiento que tiene como objetivo que un modelo que pueda clasificar objetos de clases no vistas (dominio de destino) mediante la transferencia de conocimientos obtenidos de otras clases vistas (dominio de origen) con la ayuda de información semántica (Pourpanah *et al.*, 2022).
- VOCODER, es un componente fundamental en la síntesis de voz, al proporcionar una parametrización de la forma de onda en función de parámetros específicos, que facilita su manipulación estadística y computacional (Agiomyriannakis, 2015).

En esta clave interpretativa, el despliegue contemporáneo de voces sintéticas o artificiales debe comprenderse como el resultado acumulativo de casi un siglo de exploración técnica desde que las primeras pruebas que se centraron en imitar patrones del habla humana mediante la síntesis basada en reglas durante el periodo comprendido entre los años 1930 y 1980 (Jurcys *et al.*, 2024). De esta forma, dicha evolución estuvo marcada por hitos clave que transformaron la relación entre voz y tecnología:

- 1) *TTS tradicional*. Los modelos TTS constituyen la primera gran familia de tecnologías de síntesis de voz cuya finalidad es convertir texto en habla de manera inteligible y natural (Hu *et al.*, 2022). Su desarrollo inicia con sistemas paramétricos como el VODER (Dudley, 1939), que sentaron las bases para descomponer y reconstruir la señal vocal al distinguir entre características fijas (como la vibración de las cuerdas vocales y la forma

del tracto vocal) y características variables (movimientos articulatorios como labios, lengua, paladar), que definen la voz individual y el contenido expresivo del habla. Décadas después, modelos como WaveNet (Van Den Oord *et al.*, 2016), basado en técnicas de generación autoregresiva capaces de modelar distribuciones complejas análogas a las empleadas en imágenes y texto, introdujeron la posibilidad de generar audio crudo (*raw waveforms*) con un nivel de realismo sin precedentes. Posteriormente, enfoques *end-to-end* como Tacotron (Wang *et al.*, 2017) y Tacotron 2 (Shen *et al.*, 2018) consolidaron un paradigma secuencia-a-secuencia en donde a partir de caracteres, los modelos predicen espectrogramas en escala Mel<sup>2</sup> y los transforman en formas de onda mediante *vocoders* neuronales. En esta misma línea evolutiva, la introducción de la arquitectura Transformer, cuyo mecanismo de auto-atención permite construir representaciones contextuales al atender a relaciones de largo alcance entre *tokens*<sup>3</sup> (Jurafsky y Martín, 2024) reforzó la transición hacia modelos TTS más fluidos, escalables e integrados, acercando aún más la síntesis al habla humana.

- 2) *Voice conversion*. Los sistemas de voice conversion (VC) buscan transformar una señal de voz existente para conservar el contenido lingüístico; pero adoptar el timbre, estilo o identidad vocal de otra persona, sin requerir texto como mediación (Sisman, 2020). A diferencia del *voice cloning*, que parte de muestras breves de voz objetivo, la conversión opera sobre audio completo para reestilizar o reidentificar al hablante. Sus avances decisivos surgieron con modelos *autoencoder* como VQ-VAE (Van Den Oord *et al.*, 2017), que mostraron la posibilidad de aprender representaciones discretas del habla a partir de audio crudo, y con arquitecturas generativas adversariales como StarGAN-VC (Kameoka *et al.*, 2018), que habilitaron conversiones *many-to-many* sin datos paralelos ni alineación temporal. La evolución más reciente, StarGAN-v2-VC (Li *et al.*, 2020), expandió

---

<sup>2</sup> Basada en el trabajo de Stevens y Volkman, la escala de Mel es una escala no lineal utilizada para medir el tono, que se aplica al espectro con el fin de ajustarse mejor a la respuesta de frecuencia no lineal de la cóclea humana. Es una función de potencia del mapa espacio-frecuencia de la cóclea humana y es una de las medidas psicoacústicas utilizadas para comprender la percepción del sonido (Gelfand, 2017).

<sup>3</sup> Un dato que se utiliza para representar y sustituir a otro, con el fin de evitar que la información sea vista por alguien que no está autorizado para ello (Cambridge Dictionary, s. f.).

este enfoque permitiendo conversiones *any-to-many*, transformaciones estilísticas complejas y cambios entre idiomas mediante pérdidas perceptuales que refuerzan la fidelidad a la identidad vocal objetivo. En conjunto, el VC no genera voces nuevas, sino que reconfigura performativamente voces existentes, preservando el mensaje, pero que altera la identidad acústica que lo expresa.

- 3) *Voice cloning*. Los sistemas de *voice cloning* representan una tercera familia de tecnologías de síntesis de voz cuyo propósito central es replicar el timbre, estilo y rasgos expresivos de una voz humana específica, incluso cuando sólo se dispone de muy pocos segundos de audio de referencia (Kadam *et al.*, 2024). De esta forma, el modelo o clon adquiere la capacidad de reconocer e imitar las cualidades distintivas de la voz del individuo, incluidos el tono, el timbre, el ritmo y el acento. Tras el entrenamiento, el modelo puede producir una voz sintética que se asemeja mucho a la del sujeto clonado (Genelza, 2024). Modelos como SV2TTS (Jia *et al.*, 2018) inauguraron el enfoque *zero-shot*, en el que basta con un breve fragmento de voz, sin transcripción ni ajuste adicional de parámetros, para generar habla nueva que conserva la identidad vocal del hablante original. Más recientemente, modelos como VALL-E X (Zhang *et al.*, 2023) han ampliado el paradigma mediante modelos multilingües capaces de predecir *tokens* acústicos de la voz objetivo al usar como guías tanto el audio fuente como el texto destino, logrando clonaciones altamente fieles en distintos idiomas.
- 4) *AI Dubbing pipelines*. Los *AI dubbing pipelines* representan una generación de sistemas integrados que combinan, en un solo flujo, operaciones antes separadas como traducción, alineación temporal, síntesis de voz y sincronización labial con el fin de producir doblaje automatizado de alta calidad. En este terreno, modelos como FlowDubber emplean LLMs<sup>4</sup>

---

<sup>4</sup> Un modelo lingüístico es simplemente un sistema computacional que puede predecir la siguiente palabra a partir de las palabras anteriores. Es decir, dado un contexto o prefijo de palabras, un modelo lingüístico asigna una distribución de probabilidad sobre las posibles palabras siguientes. [...] Por el contrario, los modelos lingüísticos grandes pueden predecir palabras dados contextos de miles o incluso decenas de miles de palabras. La intuición fundamental de los modelos lingüísticos es que un modelo que puede predecir texto (asignando una distribución sobre las palabras siguientes) también puede utilizarse para generar texto mediante el muestreo de la distribución (Jurafsky y Martín, 2023).

para el aprendizaje semántico y técnicas de *flow matching* en la modelación acústica, y lograr doblaje con claridad sonora, correspondencia prosódica y sincronía labial consistente (Cong *et al.*, 2025). Complementariamente, herramientas como Wav2Lip aportan sincronización labial de alta precisión mediante modelos *speaker-independent*<sup>5</sup> capaces de generar movimientos bucales que se alinean con la nueva pista de audio con niveles comparables a videos reales sincronizados (Prajwal *et al.*, 2020). Más recientemente, sistemas multimodales como DeepDubber-V1 incorporan *Chain-of-Thought* visual para interpretar estilos de doblaje y atributos finos en pantalla, guiar modelos de síntesis avanzada bajo condiciones multimodales y alcanzar mejoras significativas frente al estado del arte (Zheng, 2025). Estos *pipelines* consolidan así un enfoque holístico del doblaje automático, en el que el procesamiento audiovisual se armoniza para producir resultados más naturales, coherentes y expresivamente verosímiles.

Al respecto, Kadam *et al.* apunta que, “en los últimos años, la clonación de voz ha experimentado avances extraordinarios, impulsados por la aplicación de técnicas como *deep learning*. Abriendo oportunidades innovadoras para la interacción humano-máquina, así como la personalización y creación de contenidos” (2024, p. 1). De este modo, aunque esta tecnología se emplea desde hace décadas como herramienta de apoyo en situaciones diversas como asistentes de voz, lectores de pantalla, navegación GPS, aprendizaje de idiomas y atención telefónica automatizada, entre otras (Fraganillo, 2023), usos que antes eran experimentales y periféricos se han consolidado como una tecnología con potencial disruptivo en industrias creativas como el doblaje, lo que desdibuja los límites entre interpretación humana y síntesis algorítmica.

---

<sup>5</sup> “Complex task that involves identifying spoken words or phrases from video recordings of a speaker’s facial movements. Decoding the intricate visual dynamics of a speaker’s mouth in a high-dimensional space is a significant challenge in this field. To address this challenge, researchers have employed advanced techniques that enable machines to recognize human speech through visual cues automatically” (Nemani *et al.*, 2023, p. 1).

## VI. *AI dubbing*

El despliegue de voces artificiales no sólo amplifica la capacidad técnica para generar, imitar o transformarlas, sino que supone una variedad de tensiones sobre los distintos regímenes de protección relativos a la voz, y en este caso a su proyección hacia la práctica del doblaje. En primer lugar, los sistemas TTS tradicionales generan voces genéricas sin anclaje identitario, lo que sugiere retos mínimos vinculados únicamente al estatuto autoral del texto de entrada, ya que su uso puede ser autorizado o no. No obstante, al avanzar hacia modelos más sofisticados, los sistemas de *voice conversion*, al explícitamente transformar interpretaciones humanas en otras voces y estilos, introducen retos que abarcan desde la posible creación de obras derivadas, el uso no autorizado de obras en el entrenamiento, hasta el plagio de obras en los *outputs* (Kyi *et al.*, 2025), o la imitación directa de identidades vocales preexistentes (Fraganillo, 2023).

Aún más complejos son los modelos de *voice cloning*, pues desplazan el problema hacia la apropiación algorítmica del timbre, estilo y presencia acústica de un sujeto, convirtiendo la voz en un dato biométrico susceptible de reproducción y uso sin intervención del titular. Este desanclaje del control personal enfatiza el problema central del consentimiento y la privacidad (Genelza, 2024), pues la extracción, entrenamiento o uso ulterior de la voz puede ejecutarse sin autorización válida y erosionar la autodeterminación expresiva.

Finalmente, en el extremo más integrado del espectro, los *AI dubbing pipelines* incorporan traducción, alineación, síntesis y sincronía labial en un solo flujo computacional, habilitando la generación de simulacros completos (Gilman y Willick, 2025) de interpretación en otros idiomas, capaces de reproducir ritmo, intención y estructura expresiva del doblaje original. Sin embargo, desde una óptica autoral, la automatización integral del proceso de doblaje obliga a interrogar la naturaleza jurídica del *output* resultante. Desde el prisma del derecho de autor continental, la noción de obra exige una aportación creativa humana, y la categoría de obra derivada presupone igualmente un ejercicio humano de transformación (Lipszyc, 2017). Los *AI dubbing pipelines*, sin embargo, producen resultados complejos sin que medie una intervención creativa directamente atribuible a una persona física; lo que tensiona

las categorías autorales clásicas, ya que el resultado no es la interpretación de un artista (por tanto, no puede ampararse en derechos conexos), pero tampoco califica, sin más, como obra derivada, pues carece de la impronta creativa humana que las legislaciones exigen.

En articulación con lo precedente, su capacidad para amalgamar voz, gesto y *timing* actoral no sólo incrementa los riesgos de sustitución económica en industrias como el doblaje (Thomas, 2024), sino que exacerba peligros como suplantación identitaria; más aún cuando, como advierte Bernabo (2025), estas herramientas permiten replicar voces originales o seleccionar voces artificiales afines a partir de bancos construidos mediante la cesión indefinida de la voz del intérprete, capturada y modelada a partir de breves sesiones de grabación.

A pesar de este panorama, existen autores, como es el caso de Genelza (2020), quien plantea que la clonación de voz mediante IA puede abrir cauces expresivos inéditos, habilitar formas ampliadas de participación para personas con diversas discapacidades y, en términos más amplios, contribuir a una comunidad global más accesible, inclusiva y comunicativamente interconectada. A su vez, Bhawani Niroula (2023), divisa que su incorporación en el ámbito del doblaje y la actuación vocal ha generado transformaciones significativas, ampliando las posibilidades creativas y redefiniendo dinámicas de trabajo. Esto supondría un abanico de posibilidades que resultan atractivas en la industria, como son el caso de: *a*) una creación más diversa y auténtica de personajes, al servir la IA como auxiliar en permitir caracterizaciones más complejas; *b*) optimización de flujos de trabajo y eficiencia al poder automatizar tareas, ofrecer *feedback* en tiempo real y fortalecer procesos de grabación y edición; *c*) mayor colaboración y conectividad a nivel global; *d*) evolución en la remuneración de las estructuras involucradas en el doblaje en razón de la reducción de costos de producción y la optimización de los flujos de trabajo, y *e*) coexistencia del arte y la innovación de la inteligencia artificial como herramientas para complementar el proceso creativo.

Así, y de manera fiel a la lógica hipercapitalista que impera en la industria cultural, el uso de tecnologías de inteligencia artificial en la industria audiovisual cobra una presencia cada vez más significativa, al transformar prácticas tradicionales de doblaje, narración y posproducción. Como anticipa Matamala (2019), la evolución tecnológica no sólo diversifica las modalida-

des de acceso a los contenidos, sino que también reconfigura las preferencias del público, al moldear de nuevas maneras la forma en que las audiencias consumen y perciben los productos audiovisuales.

Ejemplos recientes lo evidencian. En *Fall* (2022), se empleó *visual dubbing* de la empresa Flawless para adaptar diálogos a diversos idiomas, así como para ajustar digitalmente los movimientos faciales de los actores para sincronizar con el nuevo audio, sin necesidad de regrabar escenas ni aplicar efectos visuales costosos (Altunin, 2023). En *The Andy Warhol Diaries* (2022), se utilizó clonación de voz autorizada a través del *software* de Resemble AI para recrear la voz del artista y narrar la serie en primera persona (Resemble AI, 2022). Y en el videoclip *Wake Me Up* (2024), la voz de Aloe Blacc fue clonada en cinco idiomas distintos por el programa Reespecher, al recombinar inteligencia artificial con sincronización facial para lograr una interpretación multilingüe, sin regrabación vocal directa (Respecher, 2024). Estos casos reflejan el alcance técnico, comercial y simbólico de la síntesis vocal algorítmica en contextos de alta visibilidad.

## VII. Casos de estudio

Bajo este contexto, y en continuidad con los retos teóricos previamente expuestos, los tres casos que siguen muestran las manifestaciones materiales de dichas tensiones, evidenciando conflictos jurídicos y éticos en torno a la clonación de voz que también repercuten directamente en la industria del doblaje.

### 1. *Roadrunner: A film about Anthony Bourdain* (2021)

El documental *Roadrunner: A Film about Anthony Bourdain* (Neville, 2021) constituye uno de los primeros casos mediáticos donde la clonación de voz con herramientas de inteligencia artificial desató un debate ético-jurídico de alcance global (Coburn *et al.*, 2022). A través de la recreación sintética de tres líneas de voz del célebre chef fallecido, la obra introdujo una voz artificial entrenada con grabaciones auténticas de Bourdain, sin advertir expresamente al espectador. Aunque el hecho fue posteriormente reconocido en entrevistas,

la omisión inicial de esta práctica en los créditos y materiales promocionales generó una reacción crítica inmediata, tanto en medios especializados como entre el público general (Rosner, 2021). Este caso permite vislumbrar que el uso de la voz en entornos de producciones audiovisuales no puede reducirse a una mera función comunicativa o a un dato biométrico técnicamente replicable, sino que constituye un elemento simbólico y jurídicamente denso, portador de identidad y expresión. Su extracción, clonación, manipulación y uso automatizados, sin autorización expresa del titular o de sus herederos, terminan por erosionar los fundamentos jurídicos de la dignidad, el consentimiento, y la identidad personal.

El núcleo del problema jurídico y ético radica en la ausencia de un consentimiento informado, explícito y verificable por parte del sujeto cuya voz fue objeto de clonación. Al tratarse de una figura pública ya fallecida, la producción incurrió en un uso de su identidad vocal sin contar con una manifestación de voluntad, ni en vida ni a través de sus herederos, respecto al uso de sistemas de inteligencia artificial para fines comerciales. Más allá de los dilemas doctrinales, Roadrunner cristaliza la conversión de elementos íntimos e identitarios en meros datos manipulables, sin mecanismos suficientes de control ni transparencia. El caso no devino en litigio, pero inauguró un precedente simbólico de gran relevancia, al normalizar el uso de voces artificiales en obras documentales sin parámetros claros de licitud.

## 2. SAG-AFTRA vs. Epic Games (2025)

El sindicato de actores SAG-AFTRA (*Screen Actors Guild–American Federation of Television and Radio Artists*) presentó una queja formal por el uso de inteligencia artificial para recrear la voz del villano Darth Vader de *Star Wars* en el videojuego *Fortnite*. La queja fue dirigida contra Llama Productions, una subsidiaria de Epic Games, desarrolladora del popular juego, por haber sustituido el trabajo de actores humanos con tecnología de IA de la compañía Eleven Labs sin previo aviso ni negociación con el sindicato. En concreto, SAG-AFTRA acusa a la empresa de haber implementado cambios unilaterales en las condiciones laborales al utilizar voces generadas por IA para realizar trabajos que tradicionalmente corresponden a actores sindicalizados sin notificar al sindicato ni brindarle la oportunidad de negociar, lo que constituiría

una violación a su obligación de negociar de buena fe, conforme a lo establecido por la Junta Nacional de Relaciones Laborales de EE. UU.

En el videojuego, los jugadores pueden interactuar con el personaje de Darth Vader; combatirlo, sumarlo a su equipo y recibir líneas de diálogo, las cuales, según la queja, habrían sido generadas mediante clonación de voz por IA en lugar de ser grabadas por un actor humano, y desplazar así el trabajo de los intérpretes (McMahon, 2025). El conflicto jurídico aquí no gira en torno al consentimiento individual del intérprete original, sino en torno a los efectos colectivos y estructurales que la sustitución automatizada tiene sobre las condiciones de trabajo, la negociación colectiva y la continuidad profesional del gremio actoral. En otras palabras, se desplaza el foco desde los derechos de la personalidad hacia los derechos laborales, evidenciando cómo el *AI dubbing* puede operar suprimiendo de facto la figura del intérprete humano sin mecanismos compensatorios, de transición o consulta.

El caso ilustra cómo el uso de inteligencia artificial no sólo plantea dilemas de propiedad o consentimiento individual, sino que también funciona como catalizador de transformaciones en la organización del trabajo creativo. El uso de modelos sintéticos entrenados sobre voces humanas permite a las empresas generar interpretaciones sin retribuir a los intérpretes, sin reconocer la labor como trabajo y sin brindar condiciones mínimas de estabilidad o participación. Este tipo de desplazamiento, lejos de ser neutro o técnicamente inevitable, responde a una lógica de maximización de eficiencia y reducción de costos, cuyas consecuencias afectan directamente la posición contractual y política del artista como sujeto laboral. En ese sentido, las consignas de la controversia presentada por SAG-AFTRA revelan que el uso de programas de clonación y generación de voz es mucho más que una disputa contractual; es una expresión anticipada de los reacomodos estructurales del trabajo artístico en la era del *AI dubbing*, donde el desplazamiento del intérprete no es sólo un efecto colateral, sino una estrategia deliberada de reorganización productiva.

### 3. *Yin - Beijing Internet Court (2024)*

Un tribunal de Beijing dictó una sentencia histórica al reconocer que imitar la voz de una persona mediante inteligencia artificial sin su consen-

miento constituye una violación a sus derechos de la personalidad. El caso fue iniciado por la actriz de doblaje Yin, cuya voz fue clonada con IA y utilizada en audiolibros sin autorización. El tribunal concluyó que la empresa de medios remitió las grabaciones de Yin a la desarrolladora de *software* sin su consentimiento, y esta última utilizó IA para clonar su voz, produciendo contenidos que luego fueron puestos a la venta. El juez Zhao Ruigang afirmó que la voz generada artificialmente reproducía con alta precisión los rasgos fonéticos, la entonación y el estilo de pronunciación de la demandante, al grado de hacerla claramente identificable. Por tal razón, se determinó que la empresa de medios culturales y la desarrolladora de *software* infringieron sus derechos al procesar y comercializar su voz sin permiso, y se ordenó una indemnización de 250,000 mil yuanes (Cao, 2024).

El pronunciamiento del tribunal constituye un hito decisivo en la protección jurídica de la voz frente a los usos no autorizados de inteligencia artificial. En el presente caso, la controversia no sólo radicó en la ausencia de consentimiento, sino en la apropiación y explotación de un *output* algorítmico (el clon vocal) cuya apariencia deriva directa y exclusivamente del material expresivo original aportado por Yin. En este sentido, el fallo pone en el centro una pregunta jurídica aún sin resolver en muchos ordenamientos: ¿quién es el titular legítimo de un clon de voz? ¿El sujeto cuya identidad lo inspira? ¿La empresa que entrena el modelo? ¿El desarrollador del *software*? La sentencia del Beijing Internet Court responde con claridad: cuando un clon de voz reproduce de manera sustancial los rasgos fonéticos y expresivos de una persona real, la titularidad simbólica y jurídica del *output* permanece en esa persona, no en los agentes técnicos o comerciales que lo procesaron.

El tribunal reconoció que la voz clonada constituye una “imitación sustancial y perceptible” del atributo personal de Yin, y que su uso comercial, sin autorización previa, constituye una violación de sus derechos de la personalidad. Este razonamiento impide que el clon vocal sea tratado como una creación nueva, separada de su referente, lo cual supone un freno normativo a las prácticas de desposesión tecnológica que han comenzado a consolidarse en industrias como el doblaje, los videojuegos y la publicidad. Asimismo, la decisión amplía la noción de responsabilidad, al extenderla a toda una cadena de agentes que participaron en la producción, distribución y monetización del clon; desde la empresa que proporcionó las graba-

ciones originales, hasta los desarrolladores de la IA y la plataforma donde se alojaron los audiolibros. Así, reconoció que los algoritmos no inventan, sino que transforman expresiones humanas preexistentes.

El caso Yin aborda el fondo del problema, la lucha por el control simbólico y económico de las nuevas formas de representación y expresión digital. La titularidad del clon de voz no puede quedar en manos del desarrollador o del empresario que lo explota, cuando su legitimidad expresiva se ancla en la identidad no sustituible de quien fue, literalmente, “clonado”. No obstante, fuera de China, pocos sistemas jurídicos han clarificado esta cuestión, de modo que la propiedad sobre los modelos entrenados y sus derivados permanece difusa.

### VIII. Retos y el futuro del doblaje ante el *AI dubbing*

Los hallazgos doctrinales y empíricos convergen en al menos tres grandes retos:

- 1) La importancia y fragilidad del consentimiento, evidenciadas en Roadranner donde la voz de Anthony Bourdain se clonó sin autorización póstuma, ilustrando cómo la IA reduce la voluntad del titular a meros “datos de entrenamiento” y reflejando que la ausencia de estándares mínimos de información, autorización, revocabilidad y monitoreo favorece un modelo de “consentimiento fantasma” donde basta localizar material de archivo para legitimar la clonación, y dejar intactas las disputas éticas sobre voluntad, autenticidad y manipulación narrativa.
- 2) El desplazamiento y la precarización laboral, encarnado en la queja de SAG-AFTRA contra Epic Games al sustituir actores sindicalizados por clonación vocal, mostrando que la clonación vocal no sólo abarata costos, sino que debilita los mínimos de remuneración, y consolidan un entorno donde la interpretación humana deviene prescindible frente a *pipelines* algorítmicos aparentemente eficientes.
- 3) La indefinición sobre la titularidad del clon de voz, resuelta a favor del intérprete por el Beijing Internet Court en el caso Yin, la cual afirmó que la réplica constituye una prolongación de la voz original y, por tanto, per-

tenece moral y patrimonialmente al intérprete, marcando un precedente aislado que aún no encuentra ecos claros en la mayoría de los sistemas jurídicos occidentales.

Al respecto, Durand (2020) ya preveía que el ser humano potenciado por la era digital no es menos vulnerable al dominio de los algoritmos que el ser humano tradicional lo fue al poder de las instituciones. En la misma línea, Jurcys *et al.* (2024) advierten que las tecnologías de síntesis vocal desbordan los supuestos que sustentan la protección de la personalidad. Esta reconfiguración sociotécnica revela los límites vigentes y exige “un nuevo contrato social con la tecnología y los principios fundamentales que nos permitan desarrollar un marco regulatorio y unas normas adecuadas para abordar los retos únicos que plantea la IA de voz” (p. 6).

Al mirar al futuro, el doblaje se encamina hacia flujos de trabajo híbridos donde la síntesis algorítmica reducirá costos y hará más eficiente los procesos de grabación y edición. Sin embargo, de no articularse un estricto marco jurídico de observancia obligatoria que combine consentimiento verificable, reparto justo de regalías y responsabilidades solidarias a lo largo de la cadena técnica, la promesa de diversidad lingüística podría derivar en lo que Yanis Varoufakis (2024, p. 115) denominó “tecnofeudalismo” que consolide la desposesión creativa. Este fenómeno implicaría la concentración del control sobre voces y datos expresivos en manos de plataformas y desarrolladores tecnológicos, que operan como nuevos señores digitales, extrayendo valor simbólico y económico de múltiples intérpretes sin garantizar contraprestaciones equitativas. Bajo esta lógica, los flujos algorítmicos propios de modelos AI terminan por oscurecer el trabajo crucial de los intérpretes, permitiendo eludir el escrutinio moral y jurídico que debería protegerlos. La labor humana queda así disfrazada como un insumo no humano apropiado por las plataformas, lo que facilita su invisibilización y consolidación como un objeto técnico más dentro del aparato productivo (Thomas, 2024).

Por el contrario, si se abren rutas regulatorias y contractuales innovadoras, la inteligencia artificial podría consolidarse en una herramienta que expanda la expresividad humana y el acceso multilingüe al contenido audiovisual. En este contexto, la protección de la integridad artística impli-

ca no sólo reconocer la autoría humana, sino garantizar que las tecnologías no disuelvan la agencia del creador bajo la apariencia de innovación.

Así, frente a estos tres retos convergentes (la fragilidad del consentimiento, la precarización y desplazamiento laborales, y la incertidumbre sobre la titularidad del clon vocal), una vía preliminar de aproximación que podrían tomar los ordenamientos ante una posible ajuste es la Regla de las 3C (consentimiento, crédito y compensación) propuesta por el Cultural Intellectual Property Rights Initiative (2018). Si bien esta regla nace para proteger expresiones culturales y saberes tradicionales frente a prácticas extractivas, su lógica estructural resulta sorprendentemente afin a los dilemas que plantea la síntesis algorítmica de la voz, toda vez que en ambos contextos existen una asimetría radical de poder, un riesgo de apropiación unilateral y una tendencia a convertir identidades expresivas (manifestaciones culturales o timbres vocales) en *inputs* desanclados de su origen humano. Bajo esta clave, el consentimiento exige procesos claros, informados, revocables y verificables que superen el modelo opaco del “consentimiento fantasma”; el crédito obliga a reconocer la procedencia humana y la agencia creativa que subsiste incluso cuando la interpretación se desmaterializa en datos, y la compensación reintroduce una dimensión distributiva indispensable para evitar que la automatización socave todavía más la posición negociadora y el sustento económico de trabajadores del doblaje (Kyi *et al.*, 2025).

Permitir que la identidad vocal se degrade a un mero recurso computacional implica despojar a los intérpretes de su agencia creativa (Thomas, 2024). Sobre esta premisa, y en consonancia con la Regla de las 3C, Jurcys *et al.* (2024) sostienen que cualquier nuevo arreglo normativo frente a la IA vocal debe partir de salvaguardar la dignidad humana, y reconocer la voz como un componente esencial de la identidad personal en el ecosistema digital. A ello se suma la necesidad de establecer estándares éticos que prioricen autorización efectiva, resguardo de la confidencialidad y deberes claros de información como condiciones mínimas para un uso justo de estas tecnologías (Genelza, 2024).

Sin embargo, sólo a través de un enfoque verdaderamente interdisciplinario y colaborativo será posible alcanzar soluciones equilibradas; normativas que protejan la dignidad de la voz, protocolos técnicos que garanticen transparencia y justicia laboral que impida la precarización masiva. Es en-

tonces imperativo emprender proyectos piloto y foros de diálogo, para que en los próximos años se pueda constituir un cuerpo de conocimiento robusto y soluciones concretas que sirvan de guía tanto a legisladores como a la industria y a los intérpretes. Estas líneas de acción no deben asumirse como restricciones al desarrollo tecnológico, sino como condiciones mínimas para una convivencia equitativa entre la innovación algorítmica y la dignidad artística, que preserve la expresividad humana como un valor irreductible, incluso en contextos profundamente mediados por la automatización.

## IX. Conclusiones

El panorama contemporáneo del doblaje revela un punto de inflexión histórico. Tras décadas de evolución técnica, la irrupción de la síntesis algorítmica de voz ha introducido un umbral inédito: el *AI dubbing*. Estas tecnologías ya no se limitan a asistir el proceso tradicional, sino que generan voces hiperrealistas capaces de modularse, clonarse y distribuirse en cuestión de segundos. Ello amplifica la promesa de accesibilidad multilingüe y eficiencia industrial, pero también reta la arquitectura jurídica que tradicionalmente amparaba la interpretación humana.

En este nuevo ecosistema, la voz se vuelve dato entrenable y reutilizable, al normalizar prácticas de extracción y replicación bajo lógicas de optimización. A partir del análisis técnico y jurídico realizado, así como del estudio de los casos *Roadrunner*, *SAG-AFTRA vs. Epic Games* y *Yin vs. Beijing Internet Court*, emergen con nitidez tres retos centrales: *a)* la fragilidad del consentimiento frente a la clonación y reutilización algorítmica de voces, incluida la posibilidad de usos póstumos o no informados; *b)* el desplazamiento y la precarización laborales derivados de la sustitución técnica de intérpretes sin mecanismos de transición, negociación ni reparto de beneficios, y *c)* la incertidumbre en torno a la titularidad simbólica y jurídica del clon vocal, especialmente cuando éste reproduce de manera sustancial la identidad acústica de una persona concreta. Todo ello revela que los marcos actuales de derechos de autor, derechos conexos, derechos de la personalidad y protección de datos biométricos ofrecen herramientas importantes, pero fragmentarias, frente a las nuevas formas de desposesión expresiva habilitadas por la síntesis

vocal algorítmica. Así, el desafío planteado por el *AI dubbing* no es meramente técnico ni estético, sino jurídico, ontológico y político.

Los hallazgos del trabajo muestran que la síntesis algorítmica vocal no sólo reconfigura procesos técnicos, sino que produce un desanclaje de la voz respecto de su corporeidad que deja al intérprete expuesto a una vulnerabilidad jurídica y laboral para la cual el derecho vigente carece de respuestas efectivas. Estos hallazgos no pretenden ofrecer un cierre definitivo ni una reforma exhaustiva del sistema, sino delinear las tensiones conceptuales que revelan la insuficiencia del andamiaje actual cuando la voz se desancla de su soporte corporal y se reconfigura como insumo algorítmico. En este contexto, la Regla de las 3C (consentimiento, crédito y compensación) y las propuestas de un “nuevo contrato social” para la voz permiten esbozar un marco mínimo de gobernanza: exigir consentimiento efectivo, revocable y verificable; garantizar reconocimiento explícito del origen humano de las voces que alimentan los modelos, y articular esquemas de compensación que eviten que la automatización se traduzca en desposesión económica y simbólica de quienes sostienen la industria del doblaje.

En suma, el artículo muestra que la voz no puede ser degradada a residuo digital ni reducida a simple recurso computacional. Incluso en su forma sintética, continúa siendo una huella de identidad que exige protección frente a dinámicas que la invisibilizan o la instrumentalizan. Reconstruir un marco de protección acorde con las condiciones del *AI dubbing* no es un lujo normativo, sino una condición mínima para que el doblaje, en la era de las voces artificiales, permanezca como espacio de mediación artística y no como laboratorio de desposesión creativa.

## X. Referencias

- Agiomyrgiannakis, Y. (2015). Vocaine: The vocoder and applications in speech synthesis. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5695-5699). IEEE. <https://ieeexplore.ieee.org/document/43336>
- Alpaydin, E. (2014). *Introduction to machine learning* (3a. ed.). MIT Press.

- Altunin, Y. (2023, enero 24). “Fall” shows us why AI-powered tools are crucial for the film industry to succeed internationally. *No Film School*. <https://nofilmschool.com/2023/01/fall-shows-us-why-ai-powered-tools-are-crucial-film-industry>
- Ávila, A. (1997). *El doblaje*. Cátedra. <https://archive.org/details/eldoblaje00avil>
- Bernabo, L. (2025). How, when, and why to use AI: Strategic uses of professional perceptions and industry lore in the dubbing industry. *International Journal of Communication*, 19, 698-715. <http://ijoc.org/>
- Bhandari, R., y Bhandari, S. (2025). Artificial intelligence: Understanding deepfakes. *EDPACS*, 61(1), 1-11. <https://doi.org/10.1080/07366981.2025.2484863>
- Biometric Information Privacy Act. (2008). 740 Ill. Comp. Stat. 14/10.
- Bosseaux, C. (2018). Dubbing. En L. Pérez-González (Ed.), *The Routledge handbook of audiovisual translation studies* (pp. 1-32). Routledge. <https://www.pure.ed.ac.uk/ws/portalfiles/portal/43181951/Bosseaux2017Dubbing.pdf>
- Cambridge Dictionary. (s. f.). Token. *Cambridge Dictionary Online*. <https://dictionary.cambridge.org/dictionary/english/token>
- Cao, Y. (2024, abril 24). AI-generated voice mimicry infringes on rights. *China Daily*. [https://english.bjinternetcourt.gov.cn/2024-04/24/c\\_706.htm](https://english.bjinternetcourt.gov.cn/2024-04/24/c_706.htm)
- Cavareto, A. (2005). *For more than one voice: Toward a philosophy of vocal expression*. Stanford University Press.
- Chaume, F. (2005). Estrategias y técnicas de traducción para el ajuste o adaptación en el doblaje. En R. Merino, J. M. Santamaría y E. Pajares (Eds.), *Trasvases culturales: literatura, cine, traducción 4* (pp. 145-153). Universidad del País Vasco. <https://www.researchgate.net/publication/361101428>
- Chaume, F. (2012). The turn of audiovisual translation: New audiences and new technologies. *Translation Spaces*, 1(2), 107-125. <https://doi.org/10.1075/ts.2.06cha>
- Chen, C. J. (2016). *Elements of human voice*. World Scientific Publishing.
- Coburn, C., Williams, K., y Stroud, S. R. (2022). Enhanced realism or A.I.-generated illusion? Synthetic voice in the documentary film *Roadrunner*. *Journal of Media Ethics*, 37(4), 1-3. <https://doi.org/10.1080/23736992.2022.2113883>

- Código Civil y Comercial de la Nación. (2015). República Argentina. <https://servicios.infoleg.gob.ar/infolegInternet/anexos/235000-239999/235975/texact.htm>
- Cong, G., Li, L., Pan, J., Zhang, Z., Beheshti, A., Van den Hengel, A., Qi, Y., y Huang, Q. (2025). FlowDubber: Movie dubbing with LLM-based semantic-aware learning and flow matching based voice enhancing. *arXiv*. <https://arxiv.org/abs/2505.01263>
- Convención de Roma. (1961). Convención sobre la protección de los artistas intérpretes o ejecutantes, los productores de fonogramas y los organismos de radiodifusión. [https://www.wipo.int/treaties/es/text.jsp?file\\_id=289795](https://www.wipo.int/treaties/es/text.jsp?file_id=289795)
- Cultural Intellectual Property Rights Initiative. (s. f.). The 3Cs Rule: Consent, Credit, Compensation. *CIPRI*. <https://www.culturalintellectualproperty.com/the-3cs>
- Dudley, H. W. (1939). *Signal transmission* (U.S. Patent No. 2,151,091). U.S. Patent and Trademark Office. <https://patents.google.com/patent/US2151091A>
- Durand, C. (2020). Tecnofeudalismo: la nueva gleba digital. *Viento Sur*, (173), 51-60. [https://biblioteca.hegoa.ehu.es/downloads/21067/%2Fsystem%2Fpdf%2F4360%2FP-VIENTO\\_SUR\\_173.5.pdf](https://biblioteca.hegoa.ehu.es/downloads/21067/%2Fsystem%2Fpdf%2F4360%2FP-VIENTO_SUR_173.5.pdf)
- Flawless AI. (2022). Fall [visual dubbing demonstration]. *Flawless AI*. <https://www.flawlessai.com>
- Fleischer, M. (2022). La precarización en la industria del doblaje. *ETERDigital*. <https://eterdigital.com.ar/la-precarizacion-en-la-industria-del-doblaje/>
- Flores Ávalos, E. L., y Pérez García, X. (2019). Protección al derecho a la imagen y a la voz ante las tecnologías de la información y comunicación. *Estudios en Derecho a la Información*, (7), e13015. <https://revistas.juridicas.unam.mx/index.php/derecho-informacion/article/view/13015/14550>
- Franganillo, J. (2023). La inteligencia artificial generativa y su impacto en la creación de contenidos mediáticos. *Methados. Revista de Ciencias Sociales*, 11(2), m231102a10. <https://doi.org/10.17502/mrcs.v11i2.710>
- Fulop, S. A., y Fitz, K. (2006). A spectrogram for the twenty-first century. *Acoustics Today*, 2(3), 24-32.
- Fundación AISGE. (2024, enero 31). 5o. Informe sociolaboral de la Fundación AISGE, 2023/24: Contenidos íntegros. <https://www.aisge.es/estudio-sociolaboral-2023-de-la-fundacion-aisge-contenidos-integros>

- Gelfand, S. A. (2017). *Hearing: An introduction to psychological and physiological acoustics* (6a. ed.). CRC Press. <https://doi.org/10.1201/9781315154718>
- Genelza, G. G. (2024). A systematic literature review on AI voice cloning generator: A game-changer or a threat? *Journal of Emerging Technologies*, 4(2), e587791. <https://doi.org/10.57040/ag587791>
- Gilman, B., y Willick, D. (2025). *Generative AI: Cultural shift towards the machine* (Honors thesis, Loyola Marymount University). Digital Commons @ LMU. <https://digitalcommons.lmu.edu/honors-thesis/587>
- Goodfellow, I., Bengio, Y., y Courville, A. (2016). *Deep learning*. MIT Press. <https://www.deeplearningbook.org>
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., y Bengio, Y. (2014). Generative adversarial nets (arXiv:1406.2661v1). *arXiv*. <https://arxiv.org/abs/1406.2661>
- Hernández Reséndiz, A. E. (1996). *Problemas de derechos de autor en el doblaje de películas* (Tesis de licenciatura). Universidad Nacional Autónoma de México. <https://ru.dgb.unam.mx/bitstream/20.500.14330/TES01000242701/3/0242701.pdf>
- Hu, C., Tan, M., Yang, Z., Zhang, J., Ren, J., y Zhao, D. (2021, December). Neural Dubber: Dubbing for videos according to scripts. *Advances in Neural Information Processing Systems (NeurIPS)*, 35. <https://arxiv.org/pdf/2110.08243>
- Indira Gandhi National Open University. (s. f.). *Communication of visual perception and its media: Dubbing, subtitling and translation* (Unit 10, Module MTT-16: Translation Studies). IGNOU. <https://egyankosh.ac.in/bitstream/123456789/93707/1/Unit-10.pdf>
- Ivarsson, J. (1998). *Subtitling*. TransEdit HB. <https://archive.org/details/subtitling0000ivar/page/n3/mode/2up>
- Jasserand, C. (2016). *Legal nature of biometric data: From “generic” personal data to sensitive data* (University of Groningen Faculty of Law Research Paper Series No. 24/2018). Universidad de Groningen. <https://ssrn.com/abstract=3230342>
- Jia, Y., Zhang, Y., Weiss, R. J., Wang, Q., Shen, J., Ren, F., Chen, Z., Nguyen, P., Pang, R., Lopez Moreno, I., y Wu, Y. (2019). Transfer learning from speaker verification to multispeaker text-to-speech synthesis. *arXiv*. <https://arxiv.org/abs/1806.04558>

- Jurafsky, D., y Martin, J. H. (2024). Appendix A: Hidden Markov Models. En *Speech and language processing* (Borrador, enero 12 de 2025). <https://web.stanford.edu/~jurafsky/slp3/A.pdf>
- Jurafsky, D., y Martin, J. H. (2023). *Speech and language processing* (3a. ed.; borrador). <https://web.stanford.edu/~jurafsky/slp3/>
- Jurcys, P., Fenwick, M., y Liaudanskas, A. (2024, agosto 24). *Voice cloning in an age of generative AI: Mapping the limits of the law and principles for a new social contract with technology*. SSRN. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4850866](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4850866)
- Kadam, S., Chougule, A., Kharat, S., Chavan, P., y Patil, M. (2024). ReVoice: A neural network based voice cloning system. En *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)* (pp. 1-6). IEEE. <https://ieeexplore.ieee.org/document/10543448>
- Kameoka, H., Kaneko, T., Tanaka, K., y Hojo, N. (2018). StarGAN-VC: Non-parallel many-to-many voice conversion with star generative adversarial networks. *arXiv*. <https://arxiv.org/abs/1806.02169>
- Kyi, L., Mahuli, A., Silberman, M. S., Binns, R., Zhao, J., y Biega, A. J. (2025). Governance of generative AI in creative work: Consent, credit, compensation, and beyond. En *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'25)*. ACM.
- Ley núm. 19.628 sobre protección de la vida privada. (2021). República de Chile. <https://www.bcn.cl/leychile/navegar?idNorma=141599>
- Ley Orgánica 1/1982, de 5 de mayo, de protección civil del derecho al honor, a la intimidad personal y familiar y a la propia imagen (1982). *Boletín Oficial del Estado*. <https://www.boe.es/buscar/act.php?id=BOE-A-1982-11196>
- Li, Y. A., Zare, A., y Mesgarani, N. (2021). StarGANv2-VC: A diverse, unsupervised, non-parallel framework for natural-sounding voice conversion. *arXiv*. <https://arxiv.org/abs/2107.10934>
- Luyken, G.-M. (1991). Overcoming language barriers in television: Dubbing and subtitling for the European audience. European Institute for the Media. <https://archive.org/details/overcominglangua0000luyk/page/74/mode/2up>
- Lipszyc, D. (2017). *Derecho de autor y derechos conexos* (Edición digital). CERLALC.
- Martínez, X. (2004). Film dubbing: Its process and translation. En P. Orero (Ed.), *Topics in audiovisual translation* (pp. 3-8). John Benjamins Publishing

- Company. [https://api.pageplace.de/preview/DT0400.9789027295125\\_A24764306/preview-9789027295125\\_A24764306.pdf](https://api.pageplace.de/preview/DT0400.9789027295125_A24764306/preview-9789027295125_A24764306.pdf)
- Magaña Rufino, J. M. (2019). *Panorama del derecho de autor en México*. Reus.
- Martínez, X. (2004). Film dubbing. En P. Orero (Ed.), *Topics in audiovisual translation* (pp. 3-8). John Benjamins Publishing Company. <https://doi.org/10.1075/btl.56.03mar>
- Matamala, A. (2010). Translations for dubbing as dynamic texts: Strategies in film synchronisation. *Babel*, 56(2), 101-118. <https://doi.org/10.1075/babel.56.2.01mat>
- McMahon, L. (2025, mayo 20). Fortnite faces complaint from actors' union over AI Darth Vader. *BBC News*. <https://www.bbc.com/news/articles/cn844l3eznqo>
- Morgen, A. (Director). (2022). *The Andy Warhol Diaries* [Serie documental]. Netflix.
- Nájar, S. (2007). *El doblaje de voz: orígenes, personajes y empresas en México*. <https://dn721804.ca.archive.org/0/items/salvador-najar-el-doblaje-de-voz-2015-vol.-i-y-ii/Salvador%20Najar%20-%20El%20doblaje%20de%20voz%20%282015%29%20-%20Vol.%20I%20y%20II%20-%20ASIN%20B010UP2LHE%20%26%20B0116VY5IO.pdf>
- Niroula, B. (2023, septiembre 12). *The impact of AI on the future of voice acting* [preimpresión]. OSF Preprints. <https://doi.org/10.31219/osf.io/wn5y3>
- Nogueira Alcalá, H. (2007). El derecho a la propia imagen como derecho fundamental implícito: Fundamentación y caracterización. *Ius et Praxis*, 13(2), 245-285. <https://www.scielo.cl/pdf/iusetp/v13n2/art11.pdf>
- Ordellín Font, J. L. (2023). El derecho a la imagen como derecho de la personalidad. *Revista de la Escuela Federal de Formación Judicial*, (53), 238-250. <https://biblio.juridicas.unam.mx/bjv/handle/123456789/41654>
- Organización Mundial de la Propiedad Intelectual. (2016). *Principios básicos del derecho de autor y los derechos conexos* (2a. ed.). [https://www.wipo.int/edocs/pubdocs/es/wipo\\_pub\\_909\\_2016.pdf](https://www.wipo.int/edocs/pubdocs/es/wipo_pub_909_2016.pdf)
- Palencia Villa, R. M. (2000). El doblaje cinematográfico: factores de eficacia desde la recepción. *Revista Latina de Comunicación Social*, (30), 5-11. [https://ddd.uab.cat/pub/artpub/2000/167448/revlatcom\\_a2000m6n30p5.pdf](https://ddd.uab.cat/pub/artpub/2000/167448/revlatcom_a2000m6n30p5.pdf)

- Palencia Villa, R. M. (2002). *La influencia del doblaje audiovisual en la percepción de los personajes* (Tesis doctoral). Universitat Autònoma de Barcelona. <https://www.tdx.cat/bitstream/handle/10803/4105/rmpv1de6.pdf?sequence=1>
- Parra Trujillo, E. de la. (2015). *Derechos de los autores, artistas e inventores*. Instituto Nacional de Estudios Históricos de las Revoluciones de México; Universidad Nacional Autónoma de México. <https://biblio.juridicas.unam.mx/bjv/detalle-libro/4019-derechos-de-los-autores-artistas-e-inventores>
- Parssinen, K. (2007). *Multilingual text-to-speech system for mobile devices: Development and applications* (Doctoral dissertation). Tampere University of Technology.
- Pavis, M., Tulti, H., y Pye, J. (2019). *Fair Pay/Play in the UK voice-over industries: A survey of 200+ voice-overs* (Report núm. 1). The Centre for Science, Culture and the Law at Exeter (SCuLE), University of Exeter. <https://doi.org/10.5281/zenodo.3340920>
- Pizzolo, C. (2025). AI, biometric data, and the effective protection of fundamental rights in the recent ECJ case-law. *Unione Europea e Diritti*, (1), 1-13.
- Pourpanah, F., Abdar, M., Luo, Y., Zhou, X., Wang, R., Lim, C. P., Wang, X.-Z., y Wu, Q. M. J. (2022). A review of generalized zero-shot learning methods. *arXiv*. <https://arxiv.org/abs/2011.08641>
- Prajwal, K. R., Mukhopadhyay, R., Namboodiri, V. P., y Jawahar, C. V. (2020). A lip sync expert is all you need for speech to lip generation in the wild. *arXiv*. <https://arxiv.org/abs/2008.10010>
- Rangel Medina, D. (1998). Los derechos de autor. En J. C. Fernández Villavicencio (Coord.), *Temas de derecho de autor y derechos conexos* (pp. 11-29). Universidad Nacional Autónoma de México. <http://ru.juridicas.unam.mx/xmlui/handle/123456789/24887>
- Resemble AI. (2022). *Resemble AI partners with Netflix for The Andy Warhol Diaries*. <https://www.resemble.ai>
- Resemble AI. (2022, marzo 9). *How Resemble AI created Andy Warhol docu-series narration using 3 minutes of original voice recordings*. <https://www.resemble.ai/andy-warhol/>
- Respeecher. (2024, junio 12). *Respeecher's AI voice cloning enables Aloe Blacc's multilingual tribute to Avicii*. <https://www.respeecher.com/case-studies/aloe-blacc-voice-cloning-singing-multilanguage>
- Respeecher. (2024). *Wake Me Up (multilingual AI-cloned performance)* [Video musical]. <https://www.respeecher.com>

- Rosner, H. (2021, julio 17). The ethics of a deepfake Anthony Bourdain voice. *The New Yorker*. <https://www.newyorker.com/culture/annals-of-gastronomy/the-ethics-of-a-deepfake-anthony-bourdain-voice>
- Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., Zhang, Y., Wang, Y., Skerry-Ryan, R. J., Saurous, R. A., Agiomyrgiannakis, Y., y Wu, Y. (2018). Natural TTS synthesis by conditioning WaveNet on mel spectrogram predictions. *arXiv*. <https://arxiv.org/abs/1712.05884>
- Sisman, B., Yamagishi, J., King, S., y Li, H. (2020). An overview of voice conversion and its challenges: From statistical modeling to deep learning. *arXiv*. <https://arxiv.org/abs/2008.03648>
- Styger, T., y Keller, E. (1994). Formant synthesis. En E. Keller (Ed.), *Fundamentals of speech synthesis and speech recognition: Basic concepts, state of the art, and future challenges* (pp. 109-128). John Wiley. <https://www.erickeller.ch/pdf-files/Styger-Keller-94-FundVol.pdf>
- Tesis: 1a./J. 166/2023 (11a.). *Semanario Judicial de la Federación, Gaceta*, Undécima Época, 1594. <https://sjf2.scjn.gob.mx/detalle/tesis/2027523>
- Thomas, S. (2024). AI and actors: Ethical challenges, cultural narratives and industry pathways in synthetic media performance. *Emerging Media*, 2(3), 523-546. <https://doi.org/10.1177/27523543241289108>
- Toda, F. (2005). Subtitulado y doblaje: Traducción especial(izada). *Quaderns. Revista de Traducción*, (12), 119-132. [https://ddd.uab.cat/pub/quaderns/2005/quaderns\\_a2005n12p119.pdf](https://ddd.uab.cat/pub/quaderns/2005/quaderns_a2005n12p119.pdf)
- Tribunal judiciaire de Paris. (2023). *Décision núm. 21/14909*.
- Trujillo Cabrera, C. (2024). El derecho a la propia imagen (y a la voz) frente a la inteligencia artificial. *InDret*, 1(2024), 74-112. <https://raco.cat/index.php/InDret/article/view/425422/519965>
- Unión de Actores y Actrices. (s. f.). *El 73 % de los actores españoles no pueden vivir de su trabajo, según informa AISGE*. <https://www.uniondeactores.com/index.php/noticias/4629-el-73-de-los-actores-espanoles-no-pueden-vivir-de-su-trabajo-segun-informa-aisge>
- Van Den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., y Kavukcuoglu, K. (2016). WaveNet: A generative model for raw audio. *arXiv*. <https://arxiv.org/abs/1609.03499>
- Van Den Oord, A., Vinyals, O., y Kavukcuoglu, K. (2018). Neural discrete representation learning. *arXiv*. <https://arxiv.org/abs/1711.00937>

- Varoufakis, Y. (2024). *Tecnofeudalismo: El sigiloso sucesor del capitalismo* (M. Valdivieso, Trad.) [Libro electrónico]. Deusto. (Publicado originalmente en 2023). [https://ia803407.us.archive.org/26/items/tecnofeudalismo-yanis-varoufakis/Tecnofeudalismo\\_Yanis\\_Varoufakis.pdf](https://ia803407.us.archive.org/26/items/tecnofeudalismo-yanis-varoufakis/Tecnofeudalismo_Yanis_Varoufakis.pdf)
- Wang, Y., Skerry-Ryan, R. J., Stanton, D., Wu, Y., Weiss, R. J., Jaitly, N., Yang, Z., Xiao, Y., Chen, Z., Bengio, S., Le, Q., Agiomyriannakis, Y., Clark, R., y Saurous, R. A. (2017). Tacotron: Towards end-to-end speech synthesis. *arXiv*. <https://arxiv.org/abs/1703.10135>
- Zabalbeascoa, P. (1997). Dubbing and the nonverbal dimension of translation. En F. Poyatos (Ed.), *Nonverbal communication and translation* (pp. 327-342). John Benjamins Publishing Company.
- Zen, H., Oura, K., Nose, T., Yamagishi, J., Sako, S., Toda, T., Masuko, T., Black, A. W., y Tokuda, K. (2009). Recent development of the HMM-based speech synthesis system (HTS). En *Proceedings of the 2009 Asia-Pacific Signal and Information Processing Association (APSIPA)*, Sapporo, Japan. <https://www.research.ed.ac.uk/en/publications/a01bb759-bbbc-46b4-b4ab-10dd3e699ca1>
- Zhang, Z., Zhou, L., Wang, C., Chen, S., Chen, Z., Liu, Y., Liu, H., Wang, H., Li, J., Lei, H., Zhao, S., y Wei, F. (2023). Speak foreign languages with your own voice: Cross-lingual neural codec language modeling. *arXiv*. <https://arxiv.org/abs/2303.03926>
- Zheng, J., Chen, Z., Ding, C., y Di, X. (2025). DeepDubber-V1: Towards high quality and dialogue, narration, monologue adaptive movie dubbing via multi-modal chain-of-thoughts reasoning guidance. *arXiv*. <https://arxiv.org/abs/2503.23660>

## Cómo citar

### IJJ-UNAM

Buendía Romero, Hugo Ignacio, “Voces sin rostro. Los retos de la inteligencia artificial en la industria del doblaje”, *Boletín Mexicano de Derecho Comparado*, México, vol. 58, núm. 174, 2025, e20248. <https://doi.org/10.22201/ijj.24484873e.2025.174.20248>

### APA

Buendía Romero, H. I. (2025). Voces sin rostro. Los retos de la inteligencia artificial en la industria del doblaje. *Boletín Mexicano de Derecho Comparado*, 58(174), e20248. <https://doi.org/10.22201/ijj.24484873e.2025.174.20248>