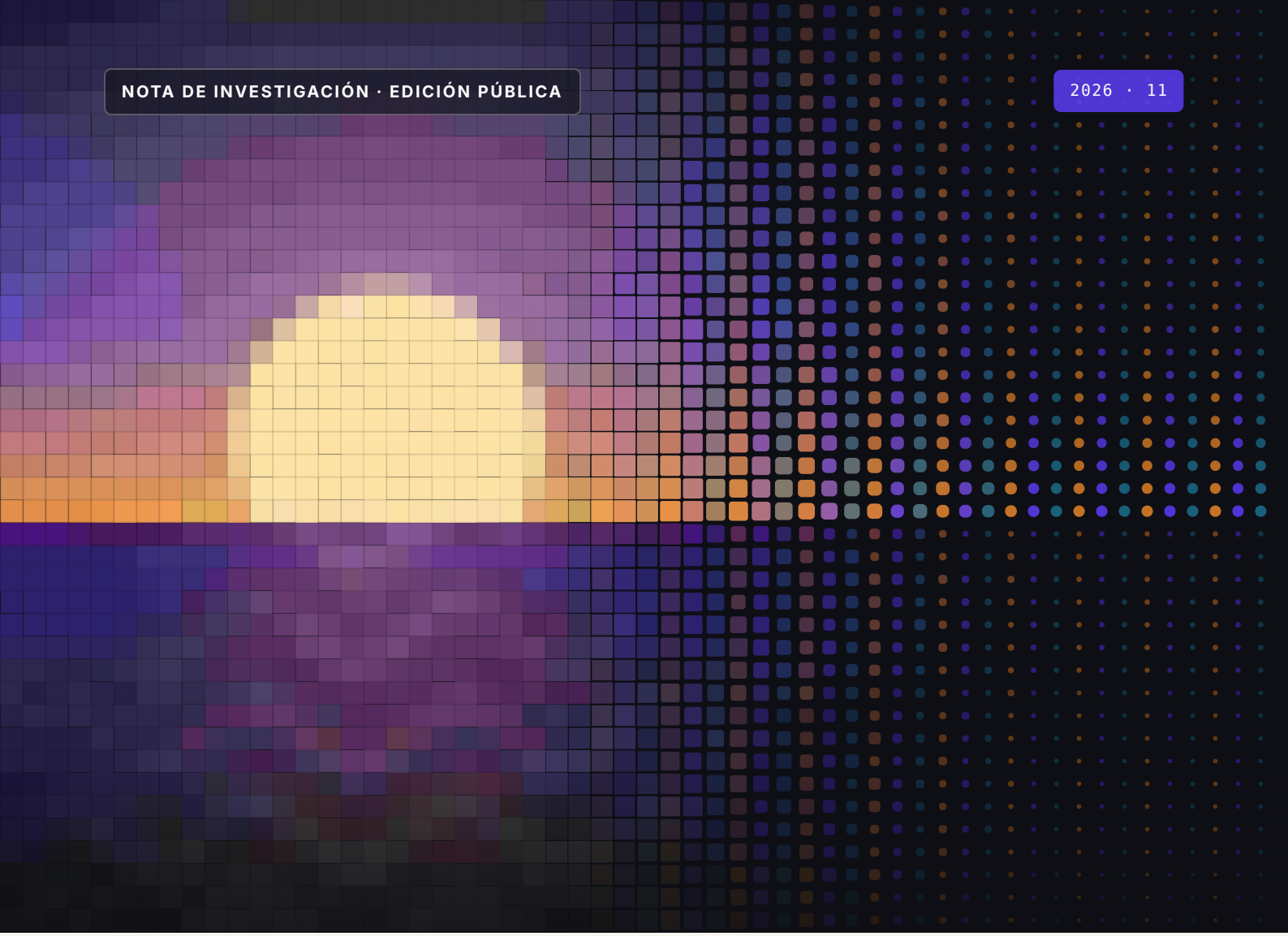


NOTA DE INVESTIGACIÓN · EDICIÓN PÚBLICA

2026 · 11

COMPRESIÓN APRENDIDA · REPRODUCCIÓN × SILICIO

Codificación neuronal en 2026

Se despliega en torno al códec, no en su lugar

Reproducción · Fotografía · Llamadas · IA generativa · Videojuegos · Silicio

> **EL VEREDICTO**
>
> **Las redes neuronales ya trabajan en ambos extremos de los códecs que transportan el vídeo del mundo. Los códecs neuronales de extremo a extremo solo se despliegan en circuitos cerrados, y las pantallas del gran público esperan a H.267.**

Tan Shuai

Tecnólogo y fundador · Investigador y desarrollador de IA

tanshuai.com · i@tanshuai.com

Datos a 12 de noviembre de 2026

Importes en dólares estadounidenses

ÍNDICE

Índice



◆ **Resumen ejecutivo** — Ya generalizada, pero en torno al códec — 04

01 **Alcance y método** — Qué significa aquí «codificación neuronal» y cómo lo estudié — 07

02 **Reproducción y streaming** — Reproducción y streaming: neuronal en ambos extremos, estándar en el medio — 09

03 **Fotografía e imágenes** — Fotografía e imágenes: neuronal dentro de la cámara, clásica en la web — 12

04 **Llamadas y reuniones** — Llamadas y reuniones: donde de verdad se despliegan los códecs neuronales de extremo a extremo — 13

05 **IA generativa** — IA generativa: el mayor despliegue que nadie llama códec — 15

06 **Videojuegos y gráficos** — Videojuegos y gráficos: la reconstrucción neuronal está generalizada, los recursos neuronales empiezan — 16

07 **Enlaces especializados** — Enlaces especializados: satélites, drones y robots — 18

08 **Silicio** — La visión desde el silicio: energía, cómputo, memoria y exactitud — 19

09 **Estándares y patentes** — Estándares y patentes: dónde las herramientas neuronales pasan a ser licenciables — 24

10 **Quién cobra** — Quién cobra y dónde puede competir una empresa — 25

11 **Afirmaciones y evidencia** — Afirmaciones habituales y lo que dice la evidencia — 27

12 **Hitos** — Qué vigilar, de 2027 a 2030 — 28

13 **Preguntas abiertas** — 29

◆ **Sobre el autor** — Tan Shuai — 30

◆ **Fuentes** — 31

FIGURAS

Figuras

FIG. 01	Dónde se despliega la codificación neuronal en 2026	06
FIG. 02	Tres formas en que una red neuronal entra en un códec	08
FIG. 03	Entre un estándar terminado y las pantallas del gran público pasa buena parte de una década	10
FIG. 04	En llamadas, un códec neuronal supera a HEVC por hardware en un 65–76 %	14
FIG. 05	Cuando una empresa controla toda la pila, la reconstrucción neuronal se extiende deprisa	17
FIG. 06	Energía por píxel decodificado: los decodificadores de función fija van un orden de magnitud por delante	20
FIG. 07	Una herramienta neuronal cuesta decenas de miles de MAC por píxel; un decodificador neuronal completo, cientos de miles	21
FIG. 08	La división que se comercializa: hardware de función fija para el flujo de bits, un motor neuronal para los píxeles	22
FIG. 09	Quién cobra por la codificación neuronal en 2026	25

Etiquetas en el texto

- **estimación** cálculo o inferencia del autor a partir de datos citados
- **dato del proveedor** publicado por la empresa que vende el producto, sin verificación independiente
- **análisis** interpretación de la evidencia por parte del autor

Traducción de la edición en inglés, que prevalece en caso de discrepancia.

Datos a 12 de noviembre de 2026 · Importes en dólares estadounidenses

RESUMEN EJECUTIVO

Ya generalizada, pero en torno al códec

EL VEREDICTO

La codificación neuronal ya está generalizada en 2026, pero funciona en torno a los códecs estándar, no en su lugar.

≈75 %
del tráfico de datos móviles es vídeo, a finales de 2025 (Ericsson)

1260 millones
de smartphones comercializados en 2025, y todos decodifican el vídeo en hardware de función fija

≈12×
la energía por píxel de un chip decodificador HEVC medido con su DRAM, para el único diseño publicado de chip decodificador neuronal

≈22 TOPS
de cómputo sostenido para un decodificador neuronal de la clase de DCVC-RT a 1080p30 estimación

75,5 %
menos tasa de bits que HEVC por hardware para MLVC a 360p en llamadas, con el propio conjunto de pruebas de Microsoft dato del proveedor

Finales de 2029
versión 1 de H.267, el primer estándar que probablemente llevará herramientas neuronales a todas las pantallas

Dónde se despliega

Las redes neuronales funcionan ya en ambos extremos de los códecs que transportan casi todo el vídeo del mundo. A la entrada, dan forma a la imagen: procesadores de señal de imagen con IA en las cámaras de los móviles y un reductor de resolución neuronal en la cadena de codificación de Netflix. A la salida, la reparan: la función Super Resolution de YouTube en los televisores, la superresolución de Meta y procesadores de televisor que anuncian cientos de redes neuronales. Entre medias, el flujo de bits sigue siendo AV1, HEVC, AVC o VVC, decodificado por hardware de función fija. Ningún servicio de streaming envía un formato de vídeo neuronal a televisores o móviles; allí donde hoy funciona un decodificador de vídeo neuronal en un dispositivo, lo hace dentro de la aplicación de una sola empresa.

Circuitos cerrados

Los códecs neuronales de extremo a extremo, en los que tanto el codificador como el decodificador son aprendidos, solo se despliegan en circuitos cerrados, donde una misma empresa controla ambos extremos y puede recurrir a un códec estándar. Microsoft usa un códec de audio con IA en todas las llamadas de Teams y Skype entre dos personas desde 2021, y en julio de 2026 liberó como código abierto MLVC, un códec de vídeo neuronal que ya se está desplegando en las llamadas entre pares de Teams: a 360p necesita alrededor de un 75 % menos de tasa de bits que HEVC por hardware para la misma puntuación de calidad. Todos los modelos de generación de voz y de vídeo funcionan sobre un códec aprendido, desde tokenizadores de voz de 12,5 Hz hasta VAE de vídeo con una relación de compresión de 1:192. La compresión neuronal de texturas ya ha llegado a un juego AAA. Son despliegues reales y comparten una condición: nadie más tiene que decodificar el flujo.

Plazos

La reproducción masiva de flujos de bits neuronales espera a un estándar y a un ciclo de silicio. El primer estándar que probablemente incluirá herramientas neuronales, el sucesor de VVC al que llamaré H.267, publicó su convocatoria de propuestas en julio de 2026, evalúa las respuestas en enero de 2027 y tiene como objetivo una primera versión a finales de 2029. AV1 necesitó casi ocho años desde la especificación terminada hasta el 30 % del visionado de Netflix. Con ese patrón, los primeros chips H.267 llegarán hacia 2031 y el uso generalizado, a mediados de la década de 2030 estimación. Entretanto, el próximo códec que llegará a las pantallas es AV2, terminado en mayo de 2026 sin ningún decodificador neuronal, y tres proveedores de IP anunciaron clientes de decodificadores AV2 en menos de tres semanas.

Silicio

La razón está en el silicio. El único diseño publicado de chip decodificador de vídeo neuronal necesita 14,6 nJ por píxel, unas 12 veces lo que necesita un chip decodificador HEVC medido, contando su DRAM, y unas 47 veces lo que necesita su núcleo solo. DCVC-RT, el códec de investigación en tiempo real de Microsoft, necesita, según mi recuento a partir de su propia tabla por módulos, unas 175.000 multiplicaciones-acumulaciones por píxel decodificado, o unos 22 TOPS de cómputo sostenido para 1080p30 estimación: alrededor de la mitad de los 40 TOPS que definen un PC con IA, dedicados a un solo flujo de vídeo. Además, un decodificador neuronal tiene que producir resultados idénticos bit a bit en todos los dispositivos, o su decodificación entrópica se desmorona. Por eso MLVC sacrifica compresión para decodificar de forma idéntica en distintas NPU y funciona a 540p, no a 4K.

Negocio

Para una empresa, el dinero sigue a los estándares y a las plataformas, no a los códecs neuronales independientes. Las plataformas se quedan el ahorro del preprocesado y el posprocesado neuronales. Los proveedores de IP licencian decodificadores de función fija para cada nuevo estándar, además de NPU para imagen. Los consorcios de patentes cobran por dispositivo: 0,50 dólares por móvil y 1,00–1,50 dólares por televisor 4K por VVC en Estados Unidos y otros mercados de la Región 1. Las tres startups de códecs neuronales más conocidas, WaveOne, iSIZE y Deep Render, las compraron Apple, Sony e InterDigital. Las oportunidades abiertas están en el preprocesado y el posprocesado neuronales, en los productos de circuito cerrado, en el silicio que combina un códec de función fija con una NPU, y en el trabajo de evaluación y estandarización que hay entre medias.

Pregunta	Juicio	Cifras clave (fuentes en los capítulos)
① **¿Dónde se despliega hoy a gran escala la codificación neuronal?**	En torno al códec: ISP de cámara, reducción de resolución al codificar, superresolución al reproducir. De extremo a extremo solo en circuitos cerrados: llamadas, recursos de juegos, tokenizadores de IA generativa	Vídeo ≈75 % de los datos móviles; 1260 millones de móviles y ≈200 millones de televisores comercializados en 2025, todos con decodificación en hardware de función fija; DLSS 4 en más de 250 juegos y aplicaciones
② **¿Sustituirán los códecs neuronales a AV1, HEVC y VVC en la reproducción? ¿Cuándo?**	No antes de un estándar y un ciclo de dispositivos. H.267 es el primer vehículo realista, y probablemente combinará herramientas clásicas y neuronales	Convocatoria de propuestas de H.267 en julio de 2026, evaluación en enero de 2027, versión 1 a finales de 2029; uso generalizado a mediados de la década de 2030 estimación
③ **¿Cuánto cuesta en silicio?**	Un orden de magnitud más de energía por píxel que la decodificación de función fija actual, más el tráfico de memoria y la exigencia de exactitud bit a bit	14,6 nJ/píxel (neuronal) frente a 1,19 (HEVC, con DRAM); ≈175 kMAC por píxel decodificado ≈ 22 TOPS a 1080p30 estimación

Pregunta	Juicio	Cifras clave (fuentes en los capítulos)
4 **¿Qué permiten los estándares y las patentes?**	Los posfiltros neuronales ya pueden señalizarse en flujos VVC; JPEG AI es un estándar ISO; H.267 decidirá qué herramientas neuronales pasan a ser licenciables	ECM (clásico) ≈28 % sobre VVC; filtros en bucle neuronales 6–14 % a 10–22+ kMAC/píxel; ByteDance, 645 familias de patentes en el trabajo «más allá de VVC»
5 **¿Quién cobra y dónde puede competir una empresa?**	Plataformas, proveedores de IP, titulares de patentes. Las startups de códecs puramente neuronales fueron adquiridas. Se puede competir en pre y posprocesado, circuitos cerrados, silicio de códec más NPU y evaluación	Consorcio de VVC: 0,50 $ por móvil; IP de decodificador AV2 licenciada a las pocas semanas de la especificación; tres adquisiciones, en condiciones no divulgadas
6 **¿Qué afirmaciones conviene tomar con cautela?**	Ganancias citadas frente a códecs antiguos, a baja resolución, sin coste de decodificación o a partir de demos en lugar de productos ya desplegados	MLVC: un 88 % menos que H.264, pero un 65–76 % menos que HEVC por hardware; texturas: ahorro de VRAM del 85 % en una demo, en torno al 30 % en el juego publicado

FIG. 01

Dónde se despliega la codificación neuronal en 2026

Las redes neuronales ya funcionan en ambos extremos de los códecs de uso general. Los códecs neuronales de extremo a extremo solo se despliegan donde una misma empresa controla el codificador y el decodificador.

	En torno a un códec estándar preprocesado, mejora, reescalado	**Códec neuronal de extremo a extremo** codificador y decodificador aprendidos	**Dentro de un estándar** herramientas neuronales en un flujo de bits o API estándar
01 **Reproducción y streaming**	● A GRAN ESCALA Reductor de resolución de Netflix; superresolución de YouTube y Meta; reescalado por IA en televisores	● SIN DESPLIEGUE Ningún servicio de streaming envía un formato de vídeo neuronal; televisores y móviles decodifican códecs estándar por hardware	● EN ESTANDARIZACIÓN La versión 1 de H.267 está prevista para finales de 2029; entre las candidatas hay herramientas neuronales
02 **Fotografía e imagen**	● A GRAN ESCALA ISP con IA en móviles de gama alta; un ISP neuronal completo, presentado como IP en el CES 2026	● DEMOS Y PILOTOS JPEG AI es un estándar ISO: demos en móviles, ningún producto	● SIN DESPLIEGUE El formato que gana terreno es clásico: JPEG XL, activado por defecto en Chrome 155
03 **Llamadas y reuniones**	● A GRAN ESCALA Superresolución en el receptor y supresión neuronal de ruido	● A GRAN ESCALA Audio neuronal en las llamadas de Teams entre dos personas desde 2021; despliegue en curso del vídeo MLVC en Teams	● DESPLIEGUE LIMITADO Opus 1.5 añade herramientas de ML opcionales y retrocompatibles
04 **IA generativa**	—	● A GRAN ESCALA Todo modelo de vídeo y de voz usa un tokenizador aprendido, p. ej., un VAE de vídeo 1:192	—
05 **Videojuegos y gráficos**	● A GRAN ESCALA Reescalado neuronal y generación de fotogramas: DLSS 4, por sí solo, está en más de 250 juegos y aplicaciones	● DESPLIEGUE LIMITADO Texturas neuronales en un juego AAA ya publicado; el SDK de NVIDIA, en beta	● EN ESTANDARIZACIÓN DirectX incorpora operaciones matriciales para shaders (versión preliminar)
06 **Enlaces especializados: satélite, robots**	● DEMOS Y PILOTOS Restauración en servidor de señales con baja tasa de bits	● DEMOS Y PILOTOS Códecs neuronales resistentes a pérdidas, en investigación y pilotos	—

Fuente: Documentos de empresas, de organismos de normalización y para desarrolladores citados en el informe; situación a 12 de noviembre de 2026. Las filas se ordenan por el número de personas al que llega cada ámbito.

ALCANCE Y MÉTODO

01 Qué significa aquí «codificación neuronal» y cómo lo estudié

Tres formas en que una red neuronal entra en un códec

«Codificación neuronal» abarca tres cosas distintas, y la mayoría de las discusiones sobre el tema nacen de confundirlas.

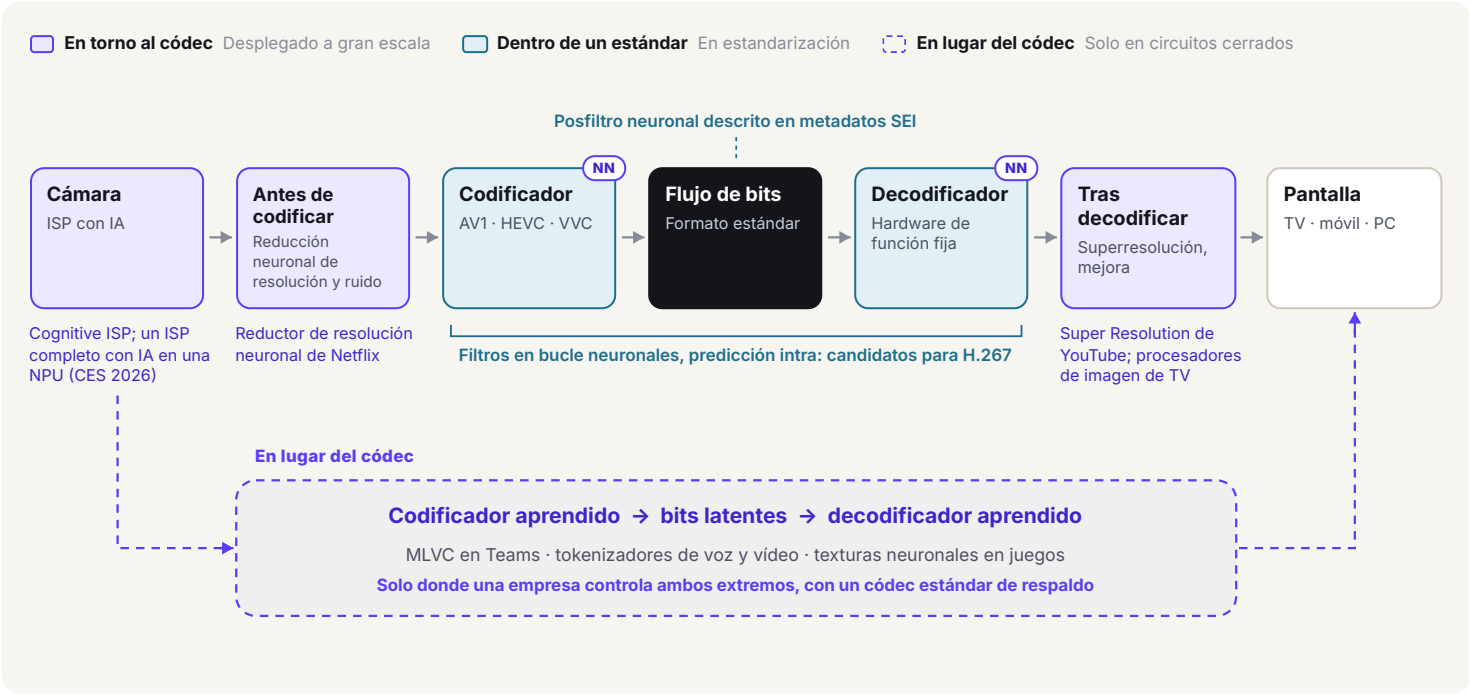
01 **En torno al códec.** Una red procesa los píxeles antes de codificar o después de decodificar, y el flujo de bits sigue siendo estándar. Ejemplos: un procesador de señal de imagen (ISP) con IA en una cámara, el reductor de resolución neuronal de Netflix antes de su codificador, la superresolución en un televisor. Cualquier dispositivo puede reproducir el flujo; la red solo cambia lo que entra o lo que sale.

02 **Dentro de un estándar.** Una red se convierte en una herramienta de un códec estándar, por ejemplo un filtro en bucle o un predictor intra neuronal que todo decodificador debe ejecutar, por lo que hay que especificarla con exactitud, normalmente en aritmética entera. Una forma más ligera es un posfiltro neuronal que el flujo de bits describe en metadatos y que los decodificadores pueden optar por aplicar.

03 **En lugar del códec.** Tanto el codificador como el decodificador se aprenden de extremo a extremo. El flujo de bits es un conjunto de valores latentes cuantificados que solo la red correspondiente puede decodificar. Ejemplos: MLVC de Microsoft en Teams, los tokenizadores de voz de los modelos de voz, las texturas neuronales de un juego.

FIG. 02 **Tres formas en que una red neuronal entra en un códec**

La mayor parte de lo que se despliega en 2026 funciona en torno a un códec estándar. Las herramientas neuronales dentro de un estándar se están especificando. Los códecs neuronales de extremo a extremo sustituyen al códec solo donde una misma empresa controla ambos extremos.



Fuente: Síntesis del autor a partir de las fuentes de este informe; situación a 12 de noviembre de 2026.

La primera forma no necesita permiso de nadie. La segunda necesita un estándar. La tercera necesita controlar ambos extremos. Ese es el orden en que llegan a la gente, y explica la mayor parte de lo que sigue.

Ordenado según el mundo real, no según la agenda de investigación

Los artículos de investigación estudian sobre todo la tercera forma, porque es donde están los récords de compresión. Este informe ordena los ámbitos según el número de personas a las que llegan, porque eso es lo que decide adónde van el dinero y la ingeniería: primero la reproducción y el streaming, después la fotografía, las llamadas, la IA generativa y los videojuegos, y en último lugar los enlaces especializados (satélites, drones, robots). El vídeo supone alrededor del 75 % de todo el tráfico de datos móviles, según el recuento de Ericsson a finales de 2025. En 2025 la industria comercializó unos 1260 millones de smartphones (IDC) y unos 196–209 millones de televisores (estimaciones de TrendForce y Omdia). Todos ellos decodifican el vídeo en hardware de función fija.

Cómo trabajé

Es una investigación documental, hecha como revisaría un chip o un plan de producto. Leí documentos de estándares e informes de reuniones (JVET, MPEG, JPEG, AOMedia, IETF), blogs de ingeniería y notas de versión de empresas, artículos sobre chips, tarifas de consorcios de patentes, textos de control de exportaciones y prensa. Cada cifra lleva su fecha. Cuando derivo yo mismo una cifra, como el cómputo por píxel a partir de las tablas de un artículo, lo digo y la marco como **estimación**. Las afirmaciones que no pude contrastar con una fuente primaria van marcadas como **dato del proveedor**. Los importes están en dólares estadounidenses. La situación es a 12 de noviembre de 2026.

Comparo lo comparable cuando puedo, y lo digo cuando no. Las ganancias de compresión significan poco sin cuatro datos: la referencia (ganancia frente a qué), la métrica (PSNR, VMAF o un panel de espectadores), el contenido y la resolución, y cuánto cuesta decodificar. La mayor parte del desacuerdo sobre los códecs neuronales desaparece en cuanto se precisan esos cuatro datos.

CAPÍTULO 02

REPRODUCCIÓN Y STREAMING

02 Reproducción y streaming: neuronal en ambos extremos, estándar en el medio

Lo que reciben los espectadores sigue siendo un flujo de bits estándar

Los mayores servicios de vídeo han completado una transición de códec y empiezan la siguiente, y ninguna incluye un decodificador neuronal. En diciembre de 2025 Netflix informó de que AV1 transportaba alrededor del 30 % de su visionado y era su segundo códec más usado, después de AVC. AV1 llegó a los televisores solo cuando llegaron los decodificadores por hardware: Netflix añadió AV1 a su certificación de dispositivos en 2019, lo emitió a televisores desde finales de 2021 y afirma que el 88 % de los dispositivos de pantalla grande presentados a certificación entre 2021 y 2025 lo admitían, casi todos desde 2023. YouTube indicó que, a 19 de marzo de 2026, más del 90 % de su catálogo, ponderado por tiempo de visualización, estaba disponible en AV1.

AV2, el sucesor de AV1 de la Alliance for Open Media, alcanzó la versión 1.0.0 el 28 de mayo de 2026, con una compresión alrededor de un 30 % mejor que la de AV1 según las mediciones objetivas de AOMedia. YouTube prevé emitir en AV2 en la web «ya a principios de 2027». El silicio llegó enseguida: Chips&Media completó un decodificador AV2 por hardware en mayo y lo licenció en junio a una gran tecnológica norteamericana sin identificar; a 16 de junio, Allegro DVT ya había entregado su IP de decodificador AV2 a tres clientes; a 10 de junio, VeriSilicon ya había entregado una IP de vídeo compatible con AV2 a varios clientes. AV2 es un códec clásico: sus ganancias proceden de mejores particiones, predicción, transformadas y filtros en bucle (uno de ellos, un filtro de detalle guiado «basado en datos»), no de una red de extremo a extremo.

VVC, el estándar vigente de la ITU y la ISO, encontró su primer hogar en radiodifusión en Brasil, cuyo sistema DTV+ (TV 3.0) entró en explotación comercial en junio de 2026, a tiempo para el Mundial. En julio de 2026 ningún televisor a la venta en Brasil llevaba un receptor DTV+ integrado; los primeros se esperan entre finales de 2026 y principios de 2027.

En el codificador: el reductor de resolución neuronal de Netflix

El ejemplo más limpio de codificación neuronal en la reproducción funciona desde 2022 y no cambia nada de lo que tiene que hacer el dispositivo del espectador. Netflix codifica cada título a varias resoluciones, y el paso de reducción de resolución que genera 1080p, 720p y 540p a partir de un máster 4K usaba tradicionalmente un filtro fijo como Lanczos. Netflix lo sustituyó por una pequeña red neuronal, un bloque de preprocesado seguido de un bloque de redimensionado, entrenada para mejorar la calidad que se ve tras un reescalador estándar en el dispositivo. Netflix la describe como una solución que se integra sin más: ningún otro cambio en el lado de la codificación ni en el del cliente, de modo que «millones de dispositivos compatibles con el streaming de Netflix se benefician automáticamente». En las pruebas de preferencia de Netflix, alrededor del 77 % de los espectadores eligió el resultado del reductor de resolución neuronal dato del proveedor. Para mantener un coste aceptable a la escala de Netflix, «bastaron unas pocas capas de red neuronal».

Ese es el patrón que hay que recordar. La red se ejecuta una vez por título en el centro de datos, donde el cómputo es barato y puede planificarse, y su resultado es un flujo AV1 o HEVC corriente.

En la pantalla: superresolución en la aplicación y en el televisor

En el otro extremo, las redes reparan lo que perdieron el códec y la fuente. El 29 de octubre de 2025 YouTube anunció Super Resolution para su aplicación de televisión: los vídeos subidos por debajo de 1080p, de 240p a 720p, se reescalan a HD con aprendizaje automático, con el 4K como objetivo declarado. Las versiones reescaladas van etiquetadas, los espectadores pueden volver al original y los creadores pueden excluirse. Meta explicó en octubre de 2025 cómo aplica superresolución de vídeo tanto en sus servidores durante la subida, en CPU x86 estándar en colaboración con Intel, como en el cliente durante la reproducción, y constató que ambas se complementan. Los fabricantes de televisores hacen lo mismo en silicio: el procesador 8K insignia de Samsung, el NQ8 AI Gen3, anuncia 768 redes neuronales para el reescalado.

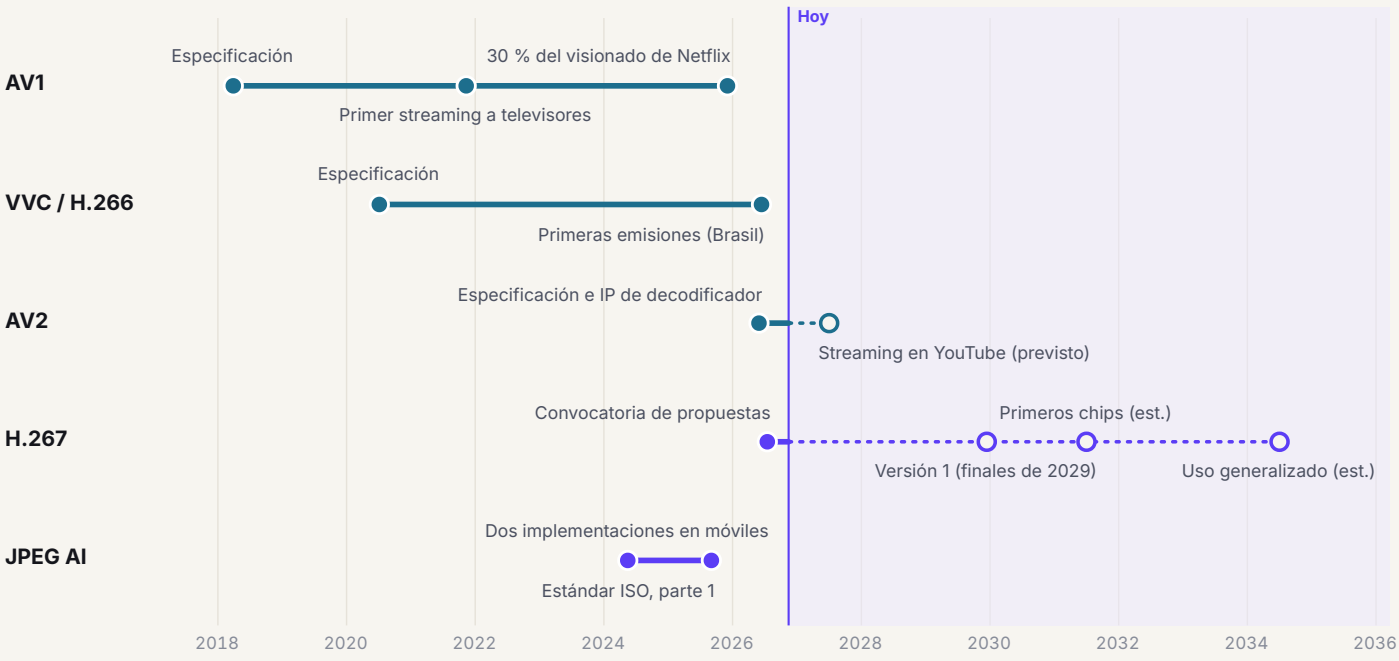
La economía cambia según dónde se ejecute la red. El procesado en el servidor le cuesta a la plataforma cómputo por cada vídeo que toca. El procesado en el dispositivo no le cuesta nada a la plataforma: usa una NPU que el comprador ya ha pagado. Por eso las mayores ganancias de 2026 proceden de ambos.

Por qué ningún televisor decodifica un flujo de bits neuronal

Un formato de reproducción tiene que decodificarse en todos los dispositivos que vayan a reproducirlo, y esos dispositivos duran años. Eso hace lento el camino de un estándar terminado a las pantallas del gran público, incluso para los códecs clásicos. La especificación de AV1 se terminó en marzo de 2018; el decodificador por software dav1d llegó a los pocos meses; los televisores lo recibieron en streaming desde finales de 2021, y el códec alcanzó alrededor del 30 % del visionado de Netflix en diciembre de 2025, casi ocho años después de la especificación. VVC se terminó en julio de 2020 y llegó a su primer servicio de radiodifusión en junio de 2026.

FIG. 03 **Entre un estándar terminado y las pantallas del gran público pasa buena parte de una década**

La reproducción necesita un decodificador por hardware en el dispositivo. AV1 tardó casi ocho años desde la especificación terminada hasta el 30 % del visionado de Netflix; el primer estándar que probablemente incluirá herramientas neuronales está previsto para finales de 2029.



Fuente: Netflix Tech Blog (dic. 2025); AOMedia; Allegro DVT; TV Technology; Fraunhofer IIS sobre el DTV+ de Brasil; MPEG-155; comité JPEG; ISO. Las marcas huecas son objetivos o estimaciones del autor.

Un flujo de bits neuronal añade dos problemas que los códecs clásicos resolvieron hace tiempo: la energía por píxel y la decodificación exacta bit a bit. Ambos se tratan en el capítulo 8. Juntos implican que ningún fabricante de televisores o de móviles pondrá un decodificador de vídeo neuronal en todos sus dispositivos antes de que un estándar le diga exactamente qué construir.

Qué lo cambiaría: H.267

El estándar que podría llevar herramientas neuronales a todas las pantallas es el sucesor de VVC, que desarrollan conjuntamente ITU-T e ISO/IEC en el Joint Video Experts Team (JVET). Su nombre no es definitivo; lo llamaré H.267, como hace buena parte del sector.

En octubre de 2025 JVET evaluó las respuestas a una convocatoria de evidencias: cinco respuestas de 16 empresas y laboratorios de investigación, con 20 configuraciones de códec probadas. El comité encontró evidencias de una compresión mejor que la de VVC y consideró «interesantes» tanto la tecnología convencional mejorada como la basada en redes neuronales. Los enfoques neuronales mostraron claras ganancias perceptuales, sobre todo en contenido 8K HDR, y ganancias menores en contenido de videojuegos. InterDigital, uno de los participantes, presentó un diseño basado en el Enhanced Compression Model (ECM) convencional y otro que superponía herramientas de IA a VVC.

La convocatoria formal de propuestas se emitió en la reunión de julio de 2026 en Ginebra. Las propuestas principales debían entregarse a finales de octubre de 2026, un paquete complementario se espera a mediados de diciembre y las respuestas se evaluarán en la reunión de JVET en Brisbane en enero de 2027. JVET prevé un primer modelo de prueba en el segundo semestre de 2027 y una primera versión del estándar antes de finales de 2029, y califica la iniciativa de «probablemente un esfuerzo mayor que nunca para poner en marcha un nuevo estándar».

La vía clásica pone el listón alto. ECM, construido con procesado de señal convencional, ahorra alrededor de un 28 % frente a VVC en configuración de acceso aleatorio. La exploración de redes neuronales de JVET recoge filtros en bucle que ahorran en torno a un 6 % con 10.000 multiplicaciones-acumulaciones por píxel, un 8 % con 22.000 y un 14 % en un punto de funcionamiento de alta complejidad. Mi lectura es que H.267 será muy probablemente un núcleo clásico con algunas herramientas neuronales allí donde los bits ahorrados justifiquen el coste del decodificador, y que la evaluación de enero de 2027 mostrará cuántas **análisis**.

Ya existe un puente. Los flujos VVC y HEVC pueden llevar información de mejora suplementaria (SEI) que describe un posfiltro neuronal para que lo aplique el lado del decodificador, y las nuevas ediciones de VVC y de su estándar de SEI (VSEI), terminadas en octubre de 2025, ampliaron esos mensajes. Un flujo puede indicar qué red ejecutar tras la decodificación sin obligar a todos los decodificadores a ejecutarla.

Qué significa para el negocio de la reproducción

Lo que hay que vender ahora es lo que funciona en torno al códec: preprocesado en la codificación para los servicios, y bloques de posprocesado y ajuste para chips de televisores y set-top boxes. No conviene montar un negocio de reproducción sobre un flujo de bits neuronal que tenga que decodificarse en dispositivos de terceros antes de la década de 2030. Para quien fabrica silicio, los ingresos a corto plazo están en los decodificadores AV2 y VVC; la parte neuronal es una ruta de NPU para la mejora de imagen junto a ellos (capítulo 8).

CAPÍTULO 03

FOTOGRAFÍA E IMÁGENES

03 Fotografía e imágenes: neuronal dentro de la cámara, clásica en la web

La cadena de la cámara ya es neuronal

Toda foto de un móvil insignia reciente ha pasado por redes neuronales antes de comprimirse. El Snapdragon 8 Gen 2 de Qualcomm, en móviles de 2023, introdujo un «Cognitive ISP» que aplica segmentación semántica en tiempo real durante la captura de fotos y vídeo y ajusta por separado caras, pelo, cielos y fondos; el Galaxy S23 Ultra de Samsung fue el primer móvil en usarlo en la cámara frontal. En enero de 2026, en el CES, Chips&Media y Visionary.ai mostraron lo que llaman el primer ISP completo basado en IA: una cadena de imagen de extremo a extremo que funciona íntegramente en una NPU, la WAVE-N de Chips&Media, en coma flotante de 16 bits, sin ISP de función fija, y que puede actualizarse de forma inalámbrica. «Las cadenas actuales acoplan bloques neuronales a un ISP de función fija. Nosotros sustituimos por completo el ISP convencional», dijo el consejero delegado de Visionary.ai. Tras la feria se ofreció a los fabricantes para su evaluación.

Después, el resultado se guarda como JPEG o HEIF, formatos clásicos que cualquier dispositivo puede abrir. El patrón es el mismo que en la reproducción: neuronal dentro de la cadena de una empresa, estándar en la frontera.

JPEG AI: un estándar neuronal terminado y aún sin producto

JPEG AI es el primer estándar de codificación de imagen aprendida de ISO/IEC e ITU-T (ISO/IEC 6048, ITU-T T.840). La parte 1, el sistema de codificación básico, se publicó en 2025. El comité JPEG informa de ganancias de compresión del 12,5 % al 27,9 % frente a la codificación intra de VVC, en un rango de complejidad del decodificador de 7000 a 216.000 multiplicaciones-acumulaciones por píxel. Su diseño tiene dos codificadores y tres decodificadores, entrenados conjuntamente, de modo que un receptor puede implementar un solo decodificador y aun así leer cualquier flujo. En mayo de 2024 el comité ya informaba de dos implementaciones en smartphones. En paralelo, el IEEE publicó en diciembre de 2024 IEEE 1857.11-2024, un estándar de codificación de imagen basada en redes neuronales diseñado para NPU.

A noviembre de 2026 no encontré ninguna cámara de móvil, aplicación de fotos ni navegador que escriba o lea JPEG AI por defecto.

Lo que adoptó la web en su lugar: JPEG XL

El formato de imagen que sí ganó terreno en 2026 es clásico. El 6 de octubre de 2026 Google activó por defecto la decodificación de JPEG XL en Chrome 155, con un nuevo decodificador con seguridad de memoria escrito en Rust, cuatro años después de retirar el soporte experimental. Safari decodifica JPEG XL desde 2023, y Mozilla avanza en la misma dirección. Los argumentos de Google a su favor son una compresión mejor que la de JPEG, la recompresión sin pérdidas de los archivos JPEG existentes, el HDR y la decodificación progresiva.

La lección es generalizable. Un formato que debe abrirse en todas partes gana con un decodificador rápido y determinista que funcione en cualquier CPU y pueda implementarse con seguridad de memoria. Un decodificador de imagen neuronal necesita una NPU o una GPU y aritmética exacta entre dispositivos. La codificación neuronal

gana dentro de la cadena propia de un dispositivo y en sistemas en los que la misma empresa codifica y decodifica, como un servicio de fotos en la nube que nunca entrega el archivo comprimido a nadie más.

Qué significa para el negocio de la imagen

El dinero está en la cadena de captura: IP de ISP con IA, NPU dimensionadas para imagen y el trabajo de ajuste que compran los fabricantes de cámaras. Para el almacenamiento y la distribución, un códec de imagen aprendido es hoy una herramienta de sistemas cerrados, no un formato por el que apostar.

CAPÍTULO 04

LLAMADAS Y REUNIONES

04 Llamadas y reuniones: donde de verdad se despliegan los códecs neuronales de extremo a extremo

El audio fue primero

Microsoft presentó Satin, un códec de voz basado en IA, en febrero de 2021, y dijo que ya se usaba en todas las llamadas de Teams y Skype entre dos personas. Satin ofrece voz en banda superancha, con hasta 16 kHz de ancho de banda de audio, a 6 kbps, mientras que Silk, el códec anterior de Skype, era de banda estrecha a esa tasa. La tasa de bits que ahorra se dedica a redundancia, de modo que las llamadas resisten mejor la pérdida de paquetes.

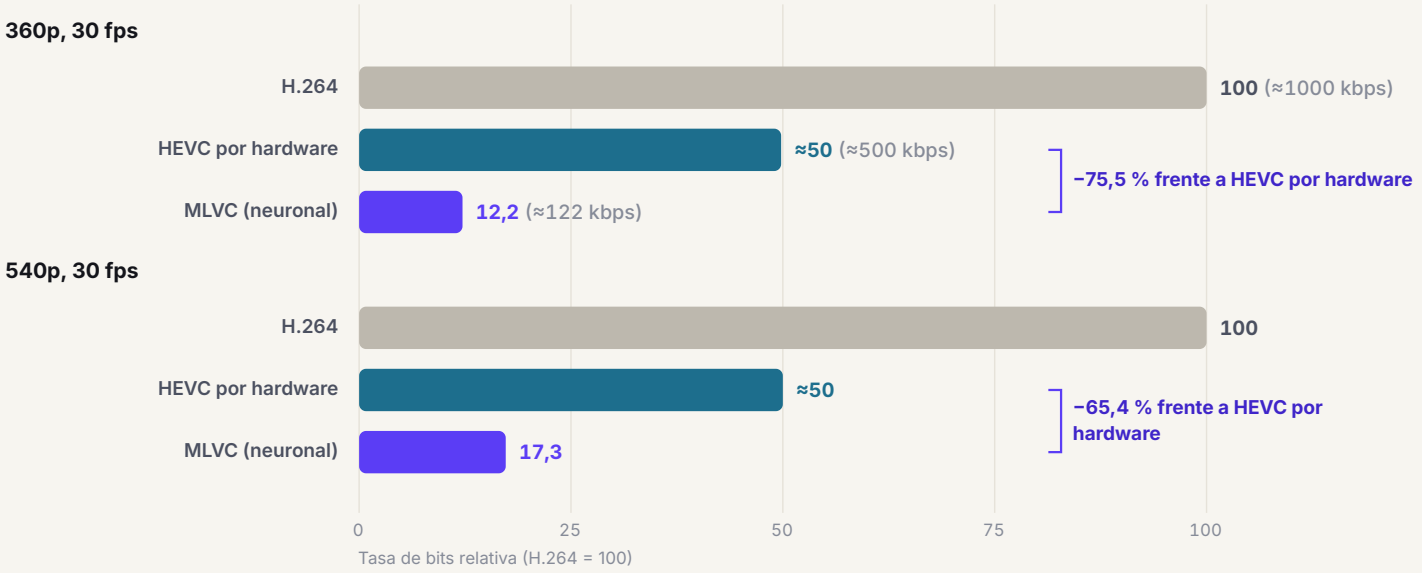
MLVC: un códec de vídeo neuronal en llamadas en producción

El 24 de julio de 2026 Microsoft publicó MLVC, su códec de vídeo basado en aprendizaje automático, como código abierto con licencia MIT, con el código del modelo, los pesos entrenados, los scripts de entrenamiento y las herramientas de conversión para distintas NPU. MLVC es la versión de producto de la familia DCVC que Microsoft Research publica desde 2021. Microsoft afirma que está desplegando MLVC en Teams para las llamadas entre pares, con telemetría en directo, pruebas A/B y los códecs convencionales como respaldo.

MLVC funciona a 540p y 360p a 30 fotogramas por segundo y está diseñado para no usar más de la mitad de la NPU en el Neural Engine de Apple y en las NPU de Qualcomm e Intel, repartiendo el trabajo entre CPU y NPU. En pruebas subjetivas según ITU-T P.910 con el Video Conferencing Dataset de la propia Microsoft, necesitó un 87,8 % menos de tasa de bits que H.264 y un 75,5 % menos que HEVC por hardware a 360p, y un 82,7 % y un 65,4 % menos a 540p **dato del proveedor**. A 360p, eso son unos 122 kbps donde H.264 necesita alrededor de 1 Mbps.

FIG. 04 **En llamadas, un códec neuronal supera a HEVC por hardware en un 65–76 %**

Tasa de bits necesaria para la misma puntuación de calidad en el conjunto de pruebas de videoconferencia de Microsoft, H.264 = 100. MLVC funciona en las NPU de portátiles y móviles hasta 540p30 y usa como máximo la mitad de la NPU.



Fuente: Microsoft, lanzamiento de MLVC (24 de julio de 2026): pruebas subjetivas según ITU-T P.910 con su Video Conferencing Dataset. Barras de HEVC derivadas de los ahorros publicados. Datos del proveedor, no mediciones independientes.

Dos cosas matizan el titular. El 88 % es frente a H.264; frente al HEVC por hardware que ya llevan esos mismos portátiles, es un 65–76 %, que sigue siendo mucho. Y MLVC renunció a compresión para poder desplegarse: un análisis independiente señala que un DCVC-RT entrenado perceptualmente obtiene un 81,9 % en la misma medida en la que MLVC obtiene un 75,5 %. El precio de decodificar de forma idéntica en hardware que no es de Microsoft fue de unos seis puntos.

Por qué las llamadas son su hogar natural

Una llamada es un circuito cerrado. Ambos extremos ejecutan la misma aplicación, negocian al inicio de la llamada lo que pueden hacer y pueden recurrir a un códec estándar en mitad de ella. Las resoluciones son modestas, las redes suelen ser malas y los portátiles de ambos extremos ya llevan NPU, como los móviles desde hace años. Microsoft fija 40 TOPS de NPU como listón para sus PC con IA, y el Neural Engine del M4 de Apple tiene una capacidad nominal de 38 TOPS. Aquí no se da ninguna de las condiciones que frenan a un códec neuronal en la reproducción.

Los contraejemplos: Opus y MLow

No todas las plataformas de llamadas tomaron la misma decisión, y los motivos son instructivos.

Opus, el códec de audio abierto de WebRTC y de la mayoría de las aplicaciones de llamadas, incorporó el aprendizaje profundo en la versión 1.5 (marzo de 2024), su primer uso de redes para procesar o generar la propia señal. Pero lo hizo en forma de opciones: una mejora en el lado del decodificador (LACE y NoLACE) que no requiere cambios en el flujo de bits, una ocultación neuronal de pérdida de paquetes y un esquema de redundancia profunda (DRED); todas están desactivadas por defecto para no sobrecargar los dispositivos antiguos. Opus 1.6 añadió una extensión neuronal del ancho de banda y un modelo DRED tres veces más pequeño, y sigue siendo plenamente compatible con el estándar de 2012. El enfoque es «dentro de un estándar, opcional y retrocompatible».

Meta fue más atrás. En junio de 2024 lanzó MLow, un códec de voz de baja tasa de bits construido sobre procesado de señal clásico en lugar de una red, que ya está en todas las llamadas de Instagram y Messenger y se está desplegando en WhatsApp. Los motivos de Meta eran sus dispositivos: más del 20 % de sus llamadas se hacen en móviles ARMv7, y decenas de millones de llamadas diarias de WhatsApp, en móviles de más de diez años. Los códecs neuronales, incluido el propio EnCodec de Meta, solo funcionan de forma fiable en terminales caros, escribió Meta. MLow ofrece cerca del doble de calidad que Opus a 6 kbps con un 10 % menos de cómputo.

Para un servicio cuyos usuarios van de portátiles de gama alta a móviles de hace una década, decide el presupuesto de complejidad. Los códecs neuronales se despliegan primero donde la base instalada es nueva y el proveedor controla ambos extremos.

Qué significa para el negocio de las comunicaciones

La videoconferencia empresarial, la telepresencia, la asistencia remota, los enlaces de contribución y las redes privadas son los lugares donde un códec neuronal de extremo a extremo es hoy un producto, siempre que ambos extremos sean propios. Conviene prever tres cosas desde el principio: un códec de respaldo, pruebas de decodificación en cada NPU compatible y una evaluación de calidad con contenido propio, no con el conjunto de datos de un proveedor.

CAPÍTULO 05

IA GENERATIVA

05 IA generativa: el mayor despliegue que nadie llama códec

Todo modelo de voz y de vídeo funciona sobre un códec aprendido

El mayor despliegue de compresión neuronal de extremo a extremo no está en ningún estándar de códec. Está dentro de la IA generativa. Un modelo de voz no escucha ni produce formas de onda directamente: lee y escribe los tokens discretos de un códec de audio aprendido. Mimi, de Kyutai, publicado en 2024, convierte la voz en tokens a 12,5 tramas por segundo y 1,1 kbps, menos que cualquier códec telefónico, y combina información semántica y acústica en un solo flujo. EnCodec, de Meta, publicado en 2022, es otro; los propios ingenieros de Meta describieron después estos códecs como excelentes a bajas tasas de bits, pero demasiado pesados para los móviles antiguos (capítulo 4).

La generación de vídeo funciona igual. Un modelo de difusión o un transformer genera en el espacio latente de un autocodificador de vídeo aprendido (VAE), y el decodificador del VAE convierte el resultado en píxeles. El VAE de LTX-Video comprime el vídeo 1:192: cada token latente representa un bloque de 32×32 píxeles a lo largo de 8 fotogramas, con 128 canales. Su decodificador también realiza el último paso de eliminación de ruido, de modo que en parte genera detalle en lugar de limitarse a reconstruirlo.

Por qué no son formatos de distribución

Estos códecs se ejecutan en cada petición a un modelo de voz o de vídeo, pero ninguno es un formato que decodifique nadie más. Cada modelo tiene su propio espacio latente; no hay control de tasa, ni prueba de conformidad, ni garantía de que otro decodificador produjera la misma imagen. Están diseñados para la calidad de generación, no para la fidelidad a una tasa de bits dada, y decodifican en GPU de centros de datos. Su éxito demues-

tra que la compresión aprendida funciona a escala cuando una misma empresa es dueña del codificador, el decodificador y el hardware.

Qué significa para las empresas

Dos consecuencias prácticas. Primera: la eficiencia del tokenizador es una partida de costes directa para quien sirve vídeo o voz generativos. En un transformer cuya atención abarca todos los tokens, el cómputo crece más deprisa que el número de tokens, así que un latente que comprime más reduce el coste de generación más que proporcionalmente análisis. Los ingenieros de códecs que entienden el diseño de latentes, los modelos de entropía y la eficiencia de los decodificadores son valiosos para los equipos de IA generativa. Segunda: la frontera se está moviendo, y los organismos de normalización han empezado a trabajar en la codificación para máquinas y no para personas. Video Coding for Machines (ISO/IEC 23888-2), de MPEG, alcanzó la fase de borrador de estándar internacional a principios de 2026 y está previsto que sea definitivo a finales de 2026, y JPEG AI se diseñó desde el principio para servir a tareas de visión artificial además de a la visualización humana.

CAPÍTULO 06

VIDEOJUEGOS Y GRÁFICOS

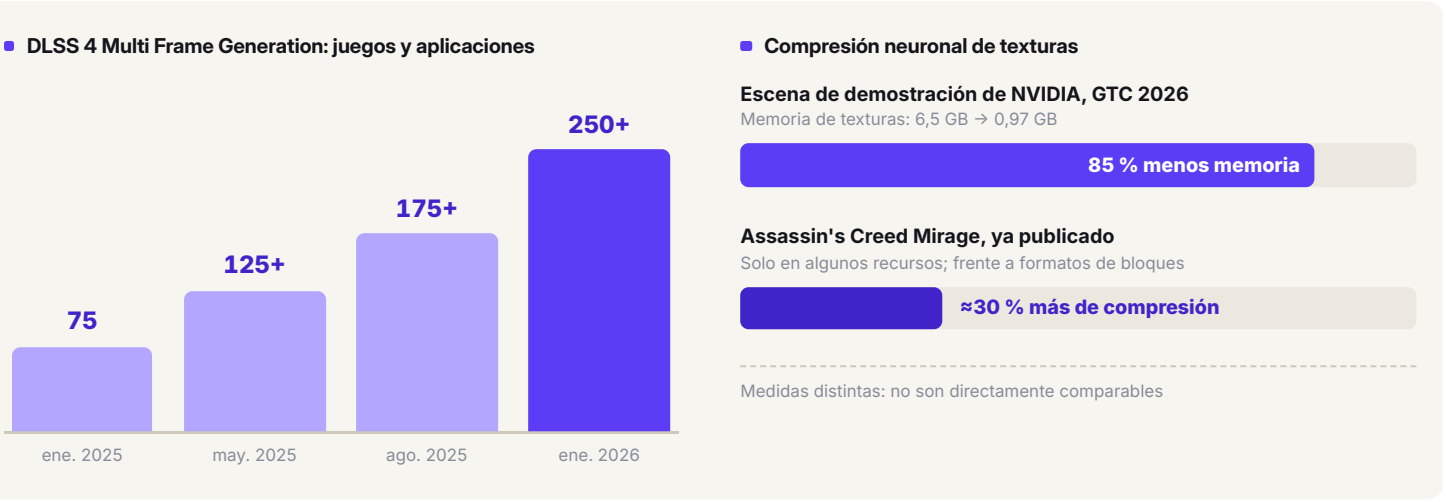
06 Videojuegos y gráficos: la reconstrucción neuronal está generalizada, los recursos neuronales empiezan

Renderizar menos píxeles y dejar que una red complete el resto

En los gráficos en tiempo real, una red ya reconstruye la mayor parte de lo que ven los jugadores. DLSS de NVIDIA renderiza menos píxeles y menos fotogramas y usa redes para producir el resto. DLSS 4 con Multi Frame Generation se lanzó en enero de 2025 con soporte en 75 juegos y aplicaciones; NVIDIA contaba más de 125 en mayo de 2025, más de 175 en agosto y más de 250 en enero de 2026, cuando anunció DLSS 4.5 con un modelo transformer de segunda generación para la superresolución. Es un decodificador en todo salvo en el nombre: el «flujo de bits» es un fotograma de baja resolución más datos de movimiento, y la red reconstruye la imagen completa.

FIG. 05 **Cuando una empresa controla toda la pila, la reconstrucción neuronal se extiende deprisa**

Izquierda: juegos y aplicaciones compatibles con DLSS 4 Multi Frame Generation de NVIDIA. Derecha: compresión neuronal de texturas en una escena de demostración y en un juego ya publicado, que solo la usó en algunos recursos por su coste en tiempo de ejecución.



Fuente: NVIDIA (de enero de 2025 a enero de 2026; demo del GTC 2026); Ubisoft La Forge (febrero de 2026). Datos del proveedor; las dos cifras de texturas miden cosas distintas.

Las texturas neuronales ya han llegado, de forma selectiva

Las texturas se comprimen en formatos de bloques fijos (BCn) que las GPU decodifican por hardware. La compresión neuronal de texturas sustituye un conjunto de texturas de material por una representación latente compacta y una pequeña red que reconstruye los téxeles cuando se necesitan. En febrero de 2026, el laboratorio La Forge de Ubisoft explicó cómo la incorporó a Assassin's Creed Mirage, un juego de 2023, con una compresión efectiva cerca de un 30 % mayor que la de los formatos convencionales con la misma calidad visual dato del proveedor. No se usó en todas partes: «su uso estuvo limitado por el rendimiento en tiempo de ejecución», así que Ubisoft la aplicó a un subconjunto de recursos con muchas instancias y una fuerte presión sobre la memoria.

El SDK RTX Neural Texture Compression de NVIDIA, en beta en GitHub, muestra el techo. En una demostración en el GTC 2026, una escena de una villa que usaba 6,5 GB de memoria de texturas con BCn necesitó 970 MB con texturas neuronales, un 85 % menos dato del proveedor. El SDK ofrece dos caminos: decodificar a BCn en el momento de la carga, lo que ahorra espacio en disco y tamaño de descarga pero no memoria de vídeo, o decodificar en el momento del muestreo, lo que ahorra la memoria pero cuesta tiempo de shader en cada fotograma.

La API está llegando

Las API gráficas están incorporando las operaciones matriciales que necesita la decodificación neuronal dentro de los shaders. Microsoft lanzó Shader Model 6.9 en febrero de 2026; la función de vectores cooperativos que presentó en versión preliminar en 2025 se integró en un diseño unificado de álgebra lineal para un shader model posterior, y DirectX Linear Algebra entró en versión preliminar pública en abril de 2026. Cuando esté en todos los controladores, un decodificador neuronal de texturas podrá ejecutarse en la GPU de cualquier fabricante desde una sola ruta de código.

Qué significa para las empresas

Los juegos despliegan decodificación neuronal porque se cumplen las condiciones de un circuito cerrado: un solo motor, una sola cadena de recursos, hardware tensorial en la GPU y un formato de respaldo. El trabajo pendiente está en las herramientas, en cadenas de recursos que codifiquen materiales neuronales, en entornos de

ejecución compatibles entre fabricantes y, sobre todo, en el móvil, donde decide el coste energético por téxel decodificado.

CAPÍTULO 07

07

ENLACES ESPECIALIZADOS

Enlaces especializados: satélites, drones y robots

Son los casos en los que la codificación neuronal de extremo a extremo encaja mejor sobre el papel. Un solo propietario controla ambos extremos del enlace, el ancho de banda es escaso y con pérdidas, el contenido es lo bastante acotado para entrenar con él, y a menudo basta con una imagen fiel en lo perceptual o incluso solo en lo semántico. Los códecs neuronales resistentes a pérdidas para estos enlaces son un área de investigación activa, y ya se utiliza la restauración en servidor de señales con baja tasa de bits.

También son un mercado pequeño y cada vez más regulado en comparación con el vídeo de consumo análisis. En China, el Anuncio n.º 49 de 2026 del Ministerio de Comercio y otros dos organismos, emitido el 19 de octubre y en vigor desde el 18 de noviembre de 2026, revisa los controles de exportación de 13 categorías de artículos relacionados con drones. Entre las entradas de tecnología, una abarca plataformas terrestres no tripuladas definidas en parte por umbrales como una compresión semántica de imagen del 75 % y la reconstrucción de la imagen en 150 milisegundos como máximo, según la prensa especializada. La propia compresión semántica se ha convertido en un parámetro de control de exportaciones. En Estados Unidos, la Comisión Federal de Comunicaciones añadió los drones de fabricación extranjera y sus componentes críticos a su Covered List el 22 de diciembre de 2025, de modo que los nuevos modelos extranjeros ya no pueden obtener la autorización de equipos.

El veredicto: un nicho sólido para una empresa especializada con clientes industriales o gubernamentales y capacidad para gestionar el cumplimiento de las normas de exportación en ambos lados, pero no un mercado masivo, ni el lugar donde averiguar si la codificación neuronal funciona a escala.

CAPÍTULO 08

SILICIO

08 La visión desde el silicio: energía, cómputo, memoria y exactitud

Este es el capítulo en el que más pesa mi propia trayectoria. Tengo detrás muchos años de I+D y de gestión de chips fabless, incluido el diseño full-custom de rutas de datos de ASIC y proyectos de chips multimedia para cámaras, como aceleradores de códecs de vídeo, y las preguntas que se hace un equipo de chips sobre un códec nuevo no tienen que ver con el BD-rate. Son estas: ¿cuántos julios por píxel?, ¿cuántos bytes de tráfico de DRAM por píxel?, ¿cuántas puertas lógicas?, ¿y pueden todos los decodificadores del mundo calcular exactamente los mismos bits? En las cuatro, un decodificador neuronal está en 2026 donde estaba HEVC por hardware antes de sus primeros chips: prometedor en software, caro en silicio.

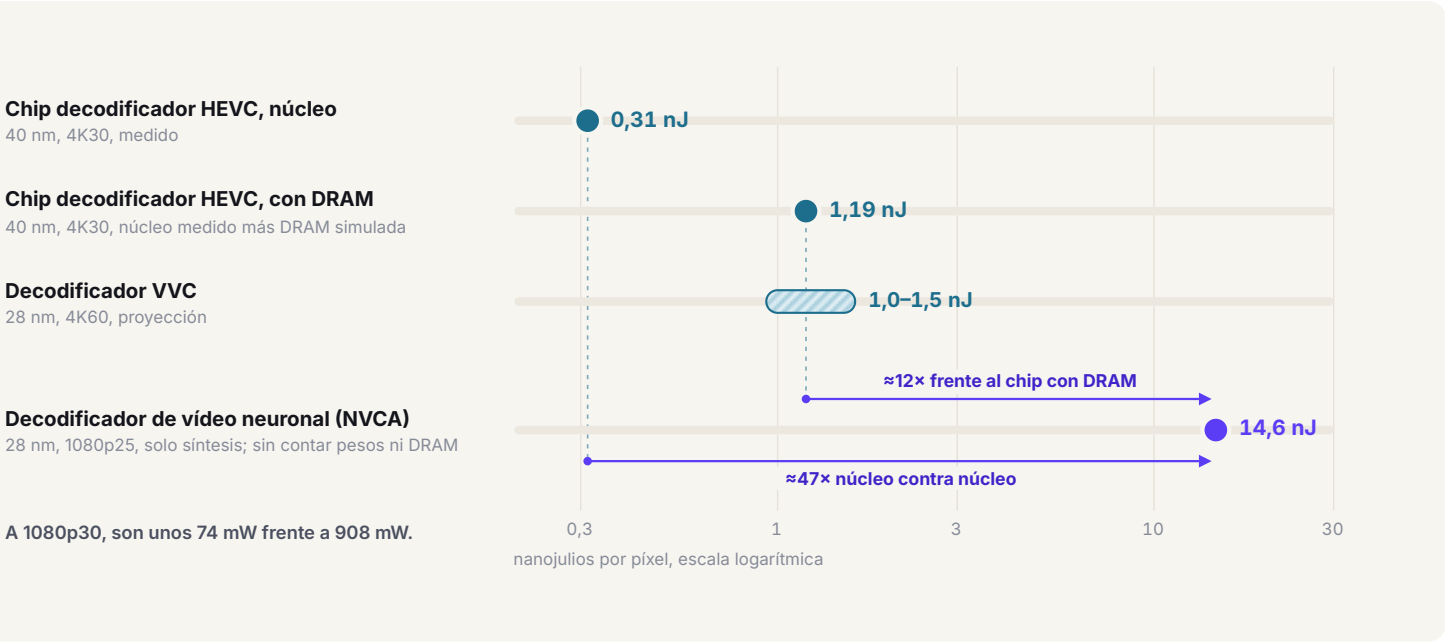
Energía por píxel: un orden de magnitud de diferencia

Pocos decodificadores neuronales se han diseñado como chips, así que la comparación se apoya en un puñado de diseños publicados. La referencia en el lado clásico es el chip decodificador HEVC del MIT (Tikekar et al., ISSCC 2013 e IEEE JSSC 2014): 40 nm, 1,77 mm², decodifica 4K a 30 fotogramas por segundo a 200 MHz. Su núcleo consume 78 mW, es decir, 0,31 nJ por píxel; contando la potencia simulada de la DRAM, el sistema necesita 1,19 nJ por píxel. Para VVC, las cifras de producción son confidenciales; un análisis publicado en Electronic Design en 2026, que escala el chip del MIT a 28 nm y aplica la mayor complejidad de VVC, proyecta 1–1,5 nJ por píxel para un decodificador 4K60 de alrededor de un millón de puertas.

En el lado neuronal, el único diseño ASIC publicado de decodificador de vídeo neuronal de extremo a extremo, según el recuento de Electronic Design, es NVCA (Zhang et al., DATE 2024), sintetizado en 28 nm de TSMC con 5,01 millones de puertas, sin contar sus pesos, que están fuera del chip. Decodifica 1080p a 25 fotogramas por segundo con 760 mW, es decir, 14,6 nJ por píxel.

FIG. 06 **Energía por píxel decodificado: los decodificadores de función fija van un orden de magnitud por delante**

El único diseño publicado de chip decodificador de vídeo neuronal necesita unas 12 veces la energía por píxel de un chip decodificador HEVC medido, contando su DRAM, y unas 47 veces la de su núcleo solo. No se cuenta el tráfico de pesos y de DRAM del propio diseño neuronal.



Fuente: Tikekar et al., IEEE JSSC 2014 (MIT); NVCA, Zhang et al., DATE 2024 (arXiv 2312.10716); el rango de VVC es una proyección de Electronic Design (2026). El chip HEVC usa un nodo más antiguo (40 nm frente a 28 nm), lo que favorece al diseño neuronal.

Eso es unas 12 veces lo que necesita el chip HEVC contando su DRAM, y unas 47 veces lo que necesita su núcleo solo. La comparación favorece a NVCA de tres maneras: el chip HEVC usa un nodo más antiguo, la cifra de NVCA deja fuera su tráfico de pesos y de DRAM, y NVCA es un resultado de síntesis, no silicio medido. A 1080p30 la diferencia es de unos 74 mW frente a cerca de 900 mW: un decodificador que nadie nota frente a otro que se nota en la batería y en el presupuesto térmico de un móvil.

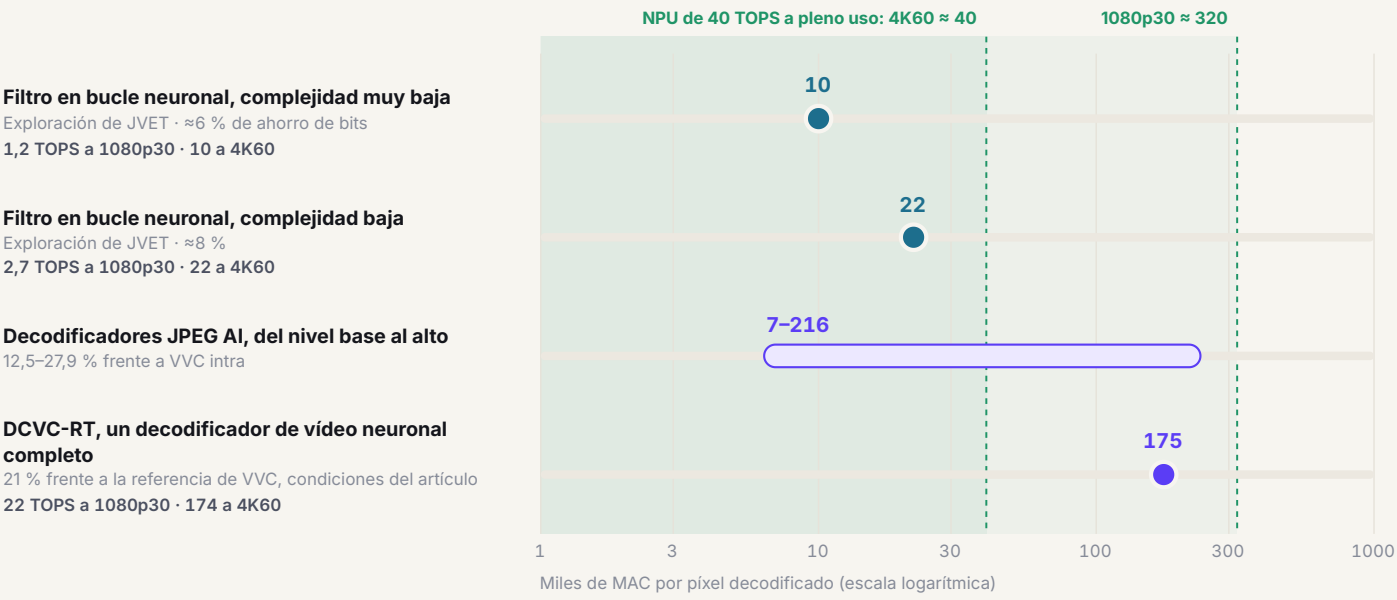
Cómputo por píxel: lo que un decodificador neuronal le pide a una NPU

El códec de vídeo neuronal publicado más rápido que encontré es DCVC-RT, de Microsoft Research (Jia et al., CVPR 2025). Codifica y decodifica 1080p a 125 y 113 fotogramas por segundo en una GPU NVIDIA A100, ahorrando un 21 % de tasa de bits frente al codificador de referencia de VVC en las condiciones de prueba del artículo, y usa aritmética entera para que distintos dispositivos decodifiquen de forma coherente. Su material complementario detalla la complejidad de cada módulo. Sumando los cuatro módulos que debe ejecutar un decodificador (decodificador, extractor de características, modelo de entropía y reconstrucción), obtengo unas 175.000 multiplicaciones-acumulaciones por píxel decodificado estimación.

A 1080p30 eso son unos 11 billones de multiplicaciones-acumulaciones por segundo, o unos 22 TOPS a dos operaciones por MAC, antes de cualquier pérdida de utilización; 4K60 necesita ocho veces más estimación. Como referencia, Microsoft define sus PC con IA por una NPU de al menos 40 TOPS, y Apple atribuye 38 TOPS al Neural Engine del M4. Son cifras de pico, compartidas con todo lo demás que hace el dispositivo.

FIG. 07 **Una herramienta neuronal cuesta decenas de miles de MAC por píxel; un decodificador neuronal completo, cientos de miles**

Multiplicaciones-acumulaciones del decodificador por píxel, en escala logarítmica, frente a lo que podría permitirse una NPU de 40 TOPS a pleno uso a 4K60 y a 1080p30. Las ganancias se miden frente a referencias distintas y no son directamente comparables.



Fuente: JVET, según el presidente del SC 29 (abril de 2026); comité JPEG (2024); DCVC-RT, material complementario de CVPR 2025, suma del autor de los módulos del decodificador; definición de PC Copilot+ de Microsoft. Cálculo del autor: dos operaciones por MAC, utilización completa.

La escalera de la figura 7 es la forma útil de leer el campo. Las herramientas neuronales que cuestan decenas de miles de MAC por píxel, como los filtros en bucle de JVET (alrededor de un 6 % de ahorro con 10.000, un 8 % con 22.000), son plausibles dentro de un futuro decodificador estándar. JPEG AI abarca todo el rango, de 7000 a 216.000 MAC por píxel, para un 12,5–27,9 % frente a VVC intra. Un decodificador de vídeo neuronal completo, con un par de cientos de miles de MAC por píxel, tiene su sitio en una NPU, a resolución modesta y en un circuito cerrado. Ahí es exactamente donde está MLVC: 540p30, usando como máximo la mitad de la NPU.

El límite real suele ser el tráfico de memoria, no los MAC

Los autores de DCVC-RT constataron que lo que limitaba la velocidad no era la aritmética sino el «coste operativo»: la entrada y salida de memoria y el número de llamadas a funciones. Eliminaron los módulos explícitos de movimiento y el procesado multiescala en gran parte por ese motivo. Los diseñadores de NVCA construyeron un flujo de datos específico para reducir el tráfico con la memoria externa de las características de movimiento y residuales. Un diseñador de chips reconocerá el patrón. En el chip HEVC del MIT, la DRAM supone unas tres cuartas partes de la energía del sistema (0,88 de 1,19 nJ por píxel), incluso después de una caché que redujo el ancho de banda de DRAM un 67 % análisis. Un decodificador neuronal que conserva grandes mapas de características multicanal entre fotogramas agrava el problema de memoria en lugar de aliviarlo.

Exactitud: todos los decodificadores deben calcular los mismos bits

La codificación entrópica solo funciona si el modelo de probabilidad del decodificador coincide exactamente con el del codificador. En un códec clásico, ese modelo es una tabla o un pequeño conjunto de reglas enteras. En un códec neuronal es una red, y una diferencia de coma flotante entre dos dispositivos, debida a otro kernel, a otro

orden de acumulación o a otra NPU, puede hacer que el decodificador lea símbolos erróneos y pierda el flujo a partir de ese punto.

Todos los que han desplegado o estandarizado un códec neuronal han tenido que resolverlo con ingeniería. DCVC-RT convierte su modelo a enteros para una codificación coherente entre dispositivos. MLVC usa un diseño de hiperprior de escala compartida para resistir las discrepancias entre plataformas, lo que probablemente explica en parte por qué cede compresión frente a DCVC-RT análisis. La exploración de redes neuronales de JVET ejecuta todas las redes con SADL, una pequeña biblioteca con inferencia en enteros de 16 bits. JPEG AI define puntos de conformidad en los dominios latente, residual y de píxeles. Para un equipo de chips es la misma disciplina que la de las transformadas enteras de HEVC: el estándar tiene que definir la aritmética con exactitud, lo que condiciona el diseño de la red y descarta atajos propios de cada fabricante.

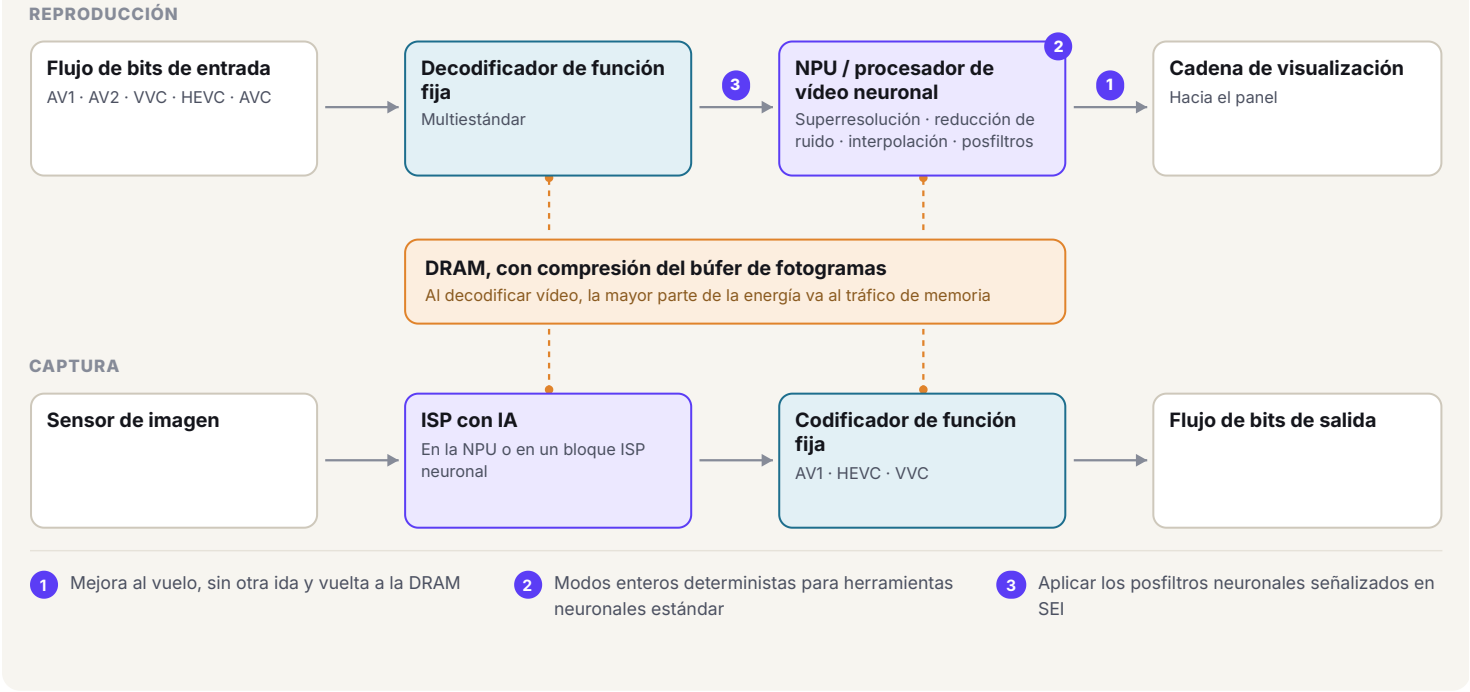
El área no es el obstáculo; la verificación y la vida útil sí

Los 5 millones de puertas de NVCA frente a alrededor de un millón de un decodificador VVC proyectado parecen muchas, pero en un SoC moderno de móvil o de televisor que ya lleva una NPU de varios TOPS, unos pocos millones de puertas no es lo que decide el diseño. Lo que cuesta dinero de verdad es la verificación frente a una batería de pruebas de conformidad, mantener correcto un decodificador programable a lo largo de las actualizaciones de software durante toda la vida útil del dispositivo, y el consumo a 4K60 HDR. Esos costes solo tienen sentido cuando un estándar fija qué hay que construir.

La arquitectura que gana hoy: un códec de función fija más una NPU

FIG. 08 **La división que se comercializa: hardware de función fija para el flujo de bits, un motor neuronal para los píxeles**

Un SoC multimedia de 2026. El bloque de códec decodifica todos los estándares con una fracción de nanojulio por píxel; la NPU hace aquello en lo que las redes destacan y que los estándares no restringen. El tráfico de DRAM es el coste que importa.



Fuente: Visión de arquitectura del autor, a partir de la información de producto publicada por Tencent (Canghai V2), Chips&Media, Allegro DVT y Samsung (2026).

El silicio que se vende en 2026 combina un códec multiestándar de función fija con un motor neuronal a su lado. El bloque de códec se encarga de AV1, AV2, VVC, HEVC y AVC con una fracción de nanojulio por píxel. El motor neuronal se encarga de aquello en lo que las redes destacan y que los estándares no restringen: la reducción de ruido y el ISP en la captura, y la superresolución, la interpolación de fotogramas y la mejora de imagen en la visualización. Ambos comparten la compresión del búfer de fotogramas y la planificación de la DRAM, porque el tráfico de memoria es el coste que importa.

El mercado refleja esa división. El ASIC de códec de vídeo Canghai (沧海) de Tencent, desarrollado internamente desde 2019, tiene más de 100.000 chips de primera generación desplegados en Tencent Cloud y en sus propios servicios en directo, bajo demanda y de juego en la nube; la segunda generación, anunciada en mayo de 2026 tras su primer encendido y ya en paso a producción en volumen para entrar en servicio en el segundo semestre del año, admite H.265 y H.266 en una sola arquitectura, con una compresión más de un 10 % y un 30 % mejor que la de su predecesora, alrededor del doble de rendimiento y mejora de imagen en el propio chip, como la superresolución. Chips&Media vende IP de códec y una NPU orientada a vídeo, WAVE-N, en paralelo. Allegro DVT vende decodificadores AV2 y ha añadido su primera IP de procesado de vídeo neuronal. Los procesadores de televisión de Samsung combinan decodificadores estándar con cientos de redes de reescalado.

Lo que yo construiría, por perfil

Equipos de SoC y de chips para televisores. Mantener decodificadores de función fija para cada estándar que haya que reproducir, incluido AV2 desde ya y VVC donde lo exija la radiodifusión. Dimensionar la NPU para el posprocesado a la resolución y la frecuencia de imagen de la pantalla, y diseñar la ruta del decodificador a la NPU y a la pantalla de modo que la mejora no añada una ida y vuelta por la DRAM. Admitir los SEI de posfiltro neuronal. Ofrecer modos enteros deterministas en la NPU. Tratar H.267 como un inicio de diseño de 2029 o posterior, y conservar cierta programabilidad para sus herramientas neuronales.

Proveedores de IP. Vender ya decodificadores AV2 y VVC, más las piezas que el procesado neuronal necesita a su alrededor: aceleradores de posfiltro, compresión del búfer de fotogramas y herramientas de conformidad.

Plataformas de vídeo. El preprocesado neuronal en el codificador es lo primero que se amortiza, porque se ejecuta una vez por título y no necesita nada de los dispositivos. Usar superresolución en el cliente donde los dispositivos lo permitan. Reservar los códecs neuronales de extremo a extremo para circuitos cerrados propios.

CAPÍTULO 09

ESTÁNDARES Y PATENTES

09 Estándares y patentes: dónde las herramientas neuronales pasan a ser licenciables

Los estándares en curso

Estándar	Organismo	Contenido neuronal	Situación, noviembre de 2026
H.267 (nombre provisional; más allá de VVC)	JVET (ITU-T e ISO/IEC)	Abierto. Tanto las herramientas clásicas (ECM) como las neuronales (filtros en bucle, predicción intra, superresolución) son candidatas	Convocatoria de propuestas en julio de 2026; respuestas evaluadas en enero de 2027; primer modelo de prueba en el segundo semestre de 2027; versión 1 a finales de 2029
VVC y HEVC, con VSEI	JVET	Un posfiltro neuronal puede describirse en metadatos SEI; aplicarlo es opcional para el decodificador	En vigor; mensajes ampliados en las ediciones de octubre de 2025
AV2	AOMedia	Herramientas clásicas, libre de regalías	Versión 1.0.0 el 28 de mayo de 2026; IP de decodificador licenciada desde junio de 2026; streaming web de YouTube previsto para principios de 2027
JPEG AI	JPEG (ISO/IEC e ITU-T)	Códec de imagen aprendido de extremo a extremo, para visión humana y artificial	Parte 1 publicada en 2025; una segunda versión en estudio; no se ha encontrado soporte por defecto en productos
IEEE 1857.11	IEEE	Codificación de imagen basada en redes neuronales, diseñada para NPU	Publicado en diciembre de 2024
Video Coding for Machines	MPEG	Codificación para el análisis automático en lugar de la visualización	Borrador de estándar internacional a principios de 2026; versión definitiva prevista para finales de 2026
Extensiones de ML de Opus	IETF	Redundancia profunda opcional, mejora en el lado del decodificador, extensión del ancho de banda	En libopus 1.5 y 1.6; redundancia profunda en estandarización

Destacan dos cosas. Primera: el contenido neuronal entra en los estándares por los bordes antes que por el núcleo, como posfiltros opcionales, mejora opcional y codificación de imagen, y todo ello deja intacta la ruta principal de decodificación. Segunda: el único estándar que podría poner herramientas neuronales en la ruta de decodificación de todas las pantallas, H.267, está a tres años de una primera versión.

Quién está bien situado en las patentes

Las herramientas de un estándar se convierten en patentes licenciables, así que el proceso de H.267 es también una carrera de patentes, y las herramientas neuronales forman parte de ella. El panorama de GreyB de 2025, con 3916 patentes «más allá de VVC», sitúa en cabeza a ByteDance, con 645 familias de patentes y 251 contribucio-

nes a JVET, y señala colaboraciones entre universidad e industria como la de Xiaomi y la Universidad de Xidian, con más de 20 familias. MLex informó en marzo de 2026 de que Tencent, Alibaba y ByteDance se están convirtiendo en participantes activos de los consorcios de patentes de códecs de vídeo. En mayo de 2026, Sauvegarder Investment Management compró a Alibaba 126 activos de patentes de códecs de vídeo que abarcan AV1, AV2 y Video Coding for Machines. La compra de Deep Render por InterDigital en octubre de 2025 añadió a su cartera las patentes de codificación de vídeo con IA de Deep Render.

El dinero por dispositivo se ve con facilidad en las tarifas de los consorcios. Las tarifas de Access Advance para VVC (a 1 de julio de 2026, para licenciatarios al corriente de sus obligaciones y que usan su marca) son de 0,50 dólares por teléfono móvil, tableta o portátil en los mercados de la Región 1, como Estados Unidos, la UE y Japón, y de 0,25 dólares en la Región 2; para televisores 4K, de 1,00 a 1,50 dólares según el precio en la Región 1; con un tope anual por empresa de 75 millones de dólares. AV1 y AV2 están libres de regalías por parte de los miembros de AOMedia. Contenga lo que contenga H.267, sus herramientas tendrán precio del mismo modo, y por eso las empresas con patentes de codificación neuronal contribuyen ya a JVET análisis.

CAPÍTULO 10

10 QUIÉN COBRA
Quién cobra y dónde puede competir una empresa

El dinero en 2026

FIG. 09 **Quién cobra por la codificación neuronal en 2026**

El dinero sigue a los estándares y a las plataformas. No encontré ninguna empresa independiente que viva de licenciar un códec de vídeo neuronal de extremo a extremo.

INGRESOS HOY

Plataformas

Se quedan el ahorro

Reductor de resolución neuronal de Netflix; superresolución de YouTube y Meta; Microsoft liberó MLVC con licencia MIT

INGRESOS HOY

Fabricantes de dispositivos y chips

Venden diferenciación

ISP de cámara; procesadores de reescalado para TV; chips de códec Canghai, de diseño propio de Tencent

INGRESOS HOY

Proveedores de IP

Licencian cada nuevo estándar

Clientes de decodificadores AV2 en Chips&Media, Allegro DVT y VeriSilicon a menos de tres semanas de la especificación

INGRESOS HOY

Fabricantes de GPU

Crean dependencia de su plataforma

DLSS 4 en más de 250 juegos y aplicaciones; SDK de texturas neuronales en beta

POR DISPOSITIVO

Titulares de patentes y consorcios

Cobran por dispositivo

VVC: 0,50 $ por móvil, 1,00–1,50 $ por televisor 4K; herramientas de H.267 después de 2029

SALIDA POR ADQUISICIÓN

Startups de códecs neuronales

Adquiridas

WaveOne → Apple (2023); iSIZE → Sony (2023); Deep Render → InterDigital (2025)

Fuente: Información publicada por empresas y consorcios de patentes, citada en el informe; situación a 12 de noviembre de 2026. Tarifas del consorcio: Access Advance, VVC, Región 1, licenciatario al corriente y con uso de la marca, a 1 de julio de 2026.

Las plataformas se quedan el ahorro. Netflix, YouTube, Meta, Microsoft y Tencent desarrollan internamente el procesado neuronal y se quedan el beneficio en forma de menos ancho de banda, menos almacenamiento y mejor calidad por bit. Cuando ayuda a la adopción, regalan el código: Microsoft publicó MLVC con licencia MIT.

Los fabricantes de dispositivos y de chips venden diferenciación. Los ISP de cámara y los procesadores de imagen de los televisores son donde las redes neuronales llegan a todos los usuarios, y se venden como funciones: el Cognitive ISP de Qualcomm, los procesadores de reescalado de Samsung.

Los proveedores de IP venden decodificadores para cada nuevo estándar, además de NPU. En menos de tres semanas desde la especificación de AV2, Chips&Media, Allegro DVT y VeriSilicon anunciaron clientes de decodificadores AV2. Esos mismos proveedores venden ahora bloques de procesado neuronal junto a los códecs.

Los fabricantes de GPU construyen plataformas. DLSS y la compresión neuronal de texturas convierten el hardware tensorial de NVIDIA en parte de la forma de hacer juegos.

Los titulares de patentes cobran por dispositivo, a través de consorcios para VVC y HEVC y, después de 2029, para H.267.

Las startups de códecs neuronales han sido adquiridas. WaveOne pasó a Apple en 2023, iSIZE a Sony Interactive Entertainment en 2023 y Deep Render a InterDigital en 2025, todas en condiciones no divulgadas. En cada caso, una plataforma o un licenciante compró el equipo y las patentes. No encontré en 2026 ninguna empresa independiente que viva de licenciar un códec de vídeo neuronal de extremo a extremo.

Dónde puede competir una empresa

Apuesta	Quién compra	Por qué ahora	Riesgo principal
Preprocesado neuronal en el codificador (reducción de resolución, reducción de ruido, ajuste por título)	Plataformas de streaming y de vídeo de usuarios, radiodifusores	No requiere cambios en los dispositivos; probado en Netflix	Las grandes plataformas lo desarrollan ellas mismas
IP y ajuste de mejora de imagen en la pantalla (superresolución, interpolación de fotogramas, calidad de imagen)	Fabricantes de chips para televisores, set-top boxes y móviles	Las NPU ya están en los dispositivos; diferenciación visible	Comoditización; desarrollo interno de los fabricantes de chips
Códecs neuronales en circuito cerrado (videoconferencia, operaciones remotas, enlaces privados)	Proveedores de comunicaciones empresariales, industriales y de telepresencia	MLVC demuestra que es viable, y su código es abierto	Hay que controlar ambos extremos; fragmentación de NPU
Silicio o IP de códec más NPU (AV2, VVC y posprocesado)	Fabricantes de SoC	Empieza el ciclo de AV2	Competidores consolidados
Herramientas de evaluación, conformidad e inferencia determinista	Cualquiera que despliegue herramientas neuronales	La exactitud bit a bit es un requisito estricto	Mercado pequeño, salvo en paquete
Contribuciones al estándar H.267	Licenciantes, grandes plataformas	Las respuestas se evalúan en enero de 2027	Largo, costoso e incierto

Para la mayoría de las empresas, las dos primeras filas son donde hay ingresos disponibles en 2027. La tercera es la más interesante para una empresa de producto especializada. La última exige paciencia y una estrategia de licencias.

CAPÍTULO 11

11 AFIRMACIONES Y EVIDENCIA
Afirmaciones habituales y lo que dice la evidencia

Afirmación	Lo que dice la evidencia
«Los códecs neuronales ya superan a VVC».	En el laboratorio, sí. DCVC-RT ahorra un 21 % frente al codificador de referencia de VVC a 1080p en las condiciones de su artículo; JPEG AI ahorra un 12,5–27,9 % frente a VVC intra. El decodificador cuesta unos 175.000 MAC por píxel, o la energía de una NPU, frente a una fracción de nanojulio en hardware de función fija.
«Los códecs de IA sustituirán a H.26x y AV1 en pocos años».	No en la reproducción. El primer estándar capaz de llevar herramientas neuronales apunta a finales de 2029, los chips llegarán hacia 2031 y el uso generalizado, a mediados de la década de 2030 estimación. El códec que llega ahora, AV2, es clásico.
«MLVC reduce el ancho de banda un 88 %».	Frente a H.264 a 360p, con el propio conjunto de pruebas de Microsoft. Frente a HEVC por hardware es un 75,5 % a 360p y un 65,4 % a 540p, en llamadas en las que Microsoft controla ambos extremos.
«Las texturas neuronales reducen la memoria de vídeo un 85–96 %».	En escenas de demostración. El juego ya publicado ganó cerca de un 30 % de compresión efectiva y solo usó texturas neuronales en algunos recursos, por el coste en tiempo de ejecución.
«JPEG AI es un estándar ISO, así que cámaras y navegadores lo adoptarán».	No encontré soporte por defecto en productos. El formato que adoptó la web en 2026, JPEG XL en Chrome 155, es clásico.
«La decodificación neuronal es demasiado cara para los móviles».	En parte es cierto. MLVC decodifica 540p30 con la mitad de una NPU; 4K60 con un modelo de la clase de DCVC-RT necesitaría unos 175 TOPS de cómputo sostenido estimación.
«Los enlaces de satélites y drones son el gran mercado del vídeo neuronal».	Técnicamente son lo que mejor encaja, pero es un mercado pequeño y cada vez más sometido a controles de exportación a ambos lados del Pacífico.
«Las grandes plataformas licenciarán un códec neuronal de una startup».	Desarrollaron el suyo (Netflix, Microsoft, Tencent) o compraron la startup (Apple, Sony, InterDigital).

Una prueba sencilla desmonta la mayoría de los argumentos de venta: pedir la referencia, la métrica, la resolución y el coste de decodificación. Una afirmación que no puede precisar las cuatro cosas es un resultado de investigación, no un producto.

CAPÍTULO 12

12 HITOS
Qué vigilar, de 2027 a 2030

Cuándo	Hito	Por qué importa
Enero de 2027	JVET evalúa en Brisbane las respuestas para H.267	Cuántas de las mejores respuestas dependen de herramientas neuronales, y a qué coste por píxel
Principios de 2027	Primeras emisiones web de YouTube en AV2; decodificadores AV2 en los SoC anunciados	El próximo ciclo clásico fija los presupuestos de los dispositivos hasta alrededor de 2030
Segundo semestre de 2027	Primer modelo de prueba de H.267	Si las herramientas neuronales están en el diseño central o solo en exploración
2027	MLVC más allá de las llamadas entre pares de Teams; otros proveedores adoptan el código abierto	La primera prueba de un códec neuronal abierto fuera de su propietario
2027–2028	Un primer producto que escriba o lea JPEG AI por defecto	El primer formato neuronal usado más allá de las fronteras de una empresa
2027–2028	DirectX Linear Algebra en controladores comerciales; texturas neuronales en más juegos y en el móvil	Decodificación neuronal entre fabricantes dentro de los gráficos
En curso	Presupuestos de NPU en móviles y PC	Con unos 100 TOPS sostenidos, un decodificador 1080p30 de la clase de DCVC-RT ocuparía alrededor de una quinta parte de la NPU en lugar de la mitad análisis
Finales de 2029	Versión 1 de H.267	El inicio del ciclo del silicio para las herramientas neuronales en la reproducción para el gran público

CAPÍTULO 13

13 Preguntas abiertas

PREGUNTAS ABIERTAS

Varias preguntas de este informe aún no tienen respuesta pública, y son las que me resultan más interesantes desde el punto de vista técnico.

Exactitud a escala. ¿Hasta qué punto puede acercarse un decodificador neuronal a una aritmética exacta bit a bit y solo con enteros en NPU de distintos fabricantes antes de renunciar a la mayor parte de su ventaja de compresión? La concesión de seis puntos de MLVC es un dato; un estándar necesita una respuesta general.

La memoria antes que los MAC. ¿Cuánto cuesta el posprocesado neuronal en tráfico de DRAM por píxel cuando se ejecuta sobre la marcha entre un decodificador por hardware y la pantalla, y cuánto puede reducirse su conjunto de trabajo?

El precio de una herramienta neuronal en H.267. ¿Con cuántos MAC y nanojulios por píxel deja un filtro en bucle neuronal de compensar los bits que ahorra en un decodificador 4K60?

Circuitos cerrados más allá de las llamadas. ¿Qué otros sistemas cerrados, desde el almacenamiento de fotos en la nube y el juego en streaming hasta los robots y los enlaces privados de contribución, pueden adoptar códecs neuronales de extremo a extremo antes de que llegue H.267?

Decodificadores que generan. Cuando un decodificador genera detalle en lugar de reconstruirlo, como ya hacen los VAE de vídeo, ¿cómo debe medirse y especificarse la fidelidad?

Son áreas que podemos explorar juntos en los próximos años, mediante el intercambio técnico y la colaboración técnica: i@tanshuai.com.

Limitaciones de este informe

Es una investigación documental basada en documentos públicos hasta el 12 de noviembre de 2026. Para esta edición no hice entrevistas primarias ni mediciones independientes. Varias cifras clave son afirmaciones de proveedores o de autores sobre sus propios conjuntos de pruebas, como el ahorro de tasa de bits de MLVC y las relaciones de compresión de las texturas neuronales; están etiquetadas como tales. Mis estimaciones de cómputo para DCVC-RT suman las cifras por módulo que publican sus autores y no tienen en cuenta las pérdidas de utilización, así que la carga real de la NPU será mayor. La comparación de energía se apoya en unos pocos diseños publicados en nodos de proceso distintos; las cifras de los decodificadores VVC y AV2 de producción son confidenciales. Los calendarios de los estándares son los objetivos de los propios comités. Las fechas de H.267 posteriores a 2029 son estimaciones mías a partir de las historias de AV1 y VVC.

Sobre el autor

Tan Shuai es tecnólogo y fundador, con más de veinte años de trabajo en chips, hardware y software de sistemas, y 20 patentes concedidas. Tiene detrás muchos años de I+D y de gestión de chips fabless, incluido el diseño full-custom de rutas de datos de ASIC y proyectos de chips multimedia para cámaras, como aceleradores de códecs de vídeo, que son la lente de este informe: energía por píxel, tráfico de memoria y qué permite un estándar que un decodificador fije en silicio. Su trabajo lo ha llevado por Silicon Valley, Nueva York, Chicago, Taiwán, el Área de la Gran Bahía (Shenzhen y Hong Kong), Vietnam y Malasia. Hoy trabaja como investigador y desarrollador de IA.

Debate técnico: i@tanshuai.com · tanshuai.com

Notas y aviso legal

Este informe es una investigación documental independiente basada en documentos públicos disponibles hasta el 12 de noviembre de 2026. Se ofrece solo a título informativo y no constituye asesoramiento de inversión, jurídico ni de ingeniería. Los productos, los calendarios de los estándares, las tarifas de los consorcios de patentes y las normas de exportación cambian; conviene verificar en la fuente primaria todo aquello en lo que uno se base. Las cifras marcadas como dato del proveedor, estimación o análisis deben leerse como tales. Los nombres de empresas y productos pertenecen a sus propietarios y se usan solo a efectos de identificación; ninguna empresa ha revisado ni patrocinado este informe.

© 2026 Tan Shuai. Se permite compartir este informe sin modificaciones y citando la autoría.

Fuentes

Todas consultadas entre el 3 y el 12 de noviembre de 2026. Agrupadas por capítulo.

Tráfico, dispositivos y alcance

- Ericsson, *Mobility Report*, junio de 2026 (el vídeo, alrededor del 75 % del tráfico de datos móviles a finales de 2025): https://www.ericsson.com/en/reports-and-papers/mobility-report/reports/june-2026 ; resumen: https://telecomlead.com/5g/ericsson-mobility-report-2026-global-mobile-data-traffic-to-reach-515-eb-by-2031-as-india-leads-smartphone-data-consumption-126290
- IDC, envíos mundiales de smartphones, 4.º trimestre y año completo 2025 (1260 millones), 13 de enero de 2026: https://www.idc.com/resource-center/press-releases/q425mobilephonetop5/
- TrendForce, previsión de envíos de televisores para 2025 (195,71 millones): https://iconnect007.com/article/146406/early-shipments-by-tv-brands-push-global-1h25-volume-up-2-but-raise-concerns-for-secondhalf-peak-season/146403/ein ; Omdia, previsión de televisores para 2025 (208,7 millones): https://iconnect007.com/article/145954/omdia-forecasts-flat-tv-shipments-in-2025-but-mini-led-poised-for-explosive-growth-at-829/145951/ein

Reproducción y streaming

- Netflix Technology Blog, «AV1 — Now Powering 30% of Netflix Streaming», diciembre de 2025: https://netflixtechblog.com/av1-now-powering-30-of-netflix-streaming-02f592242d80
- Netflix Technology Blog, «For your eyes only: improving Netflix video quality with neural networks», noviembre de 2022: https://netflixtechblog.com/for-your-eyes-only-improving-netflix-video-quality-with-neural-networks-5b8d032da09c
- AOMedia, AV2 Bitstream and Decoding Process Specification v1.0.0, 28 de mayo de 2026: https://av2.aomedia.org/v1.0.0/index.html ; «AV2: What the Final Specification Means for Implementers»: https://aomedia.org/blog%20posts/AV2-What-the-Final-Specification-Means-for-Implementers/
- TV Technology, «YouTube to Commence AV2 Streaming in 2027» (cuota de AV1 en el catálogo de YouTube): https://www.tvtechnology.com/platform/streaming/youtube-to-commence-av2-streaming-in-2027
- Google Research, «The in-loop filtering pipeline in AV2», 2026: https://research.google/pubs/the-in-loop-filtering-pipeline-in-av2/
- Fraunhofer IIS, «The Future of Broadcast in Brazil» (explotación comercial de DTV+, junio de 2026): https://mpegh.com/brazil/ ; CartaCapital sobre los receptores en los televisores: https://www.cartacapital.com.br/?p=628337 ; Olhar Digital sobre VVC en TV 3.0: https://olhardigital.com.br/2026/05/01/pro/em-breve-a-tv-aberta-vai-mudar-no-brasil/
- 9to5Google, «YouTube adds AI-upscaled 'Super Resolution' for low-quality videos», 29 de octubre de 2025: https://9to5google.com/2025/10/29/youtube-adds-ai-upscaled-super-resolution-for-low-quality-videos/
- Meta, «How Meta Deployed Super Resolution at Scale to Transform Video Quality», @Scale, 16 de octubre de 2025: https://atscaleconference.com/how-meta-deployed-super-resolution-at-scale-to-transform-video-quality/
- Samsung, características de los televisores 8K (NQ8 AI Gen3, 768 redes neuronales): https://www.samsung.com/us/tvs/8k-tv/highlights/

Fotografía e imágenes

- Qualcomm, «Meet the Snapdragon 8 Gen 2 Super Cameras» (Cognitive ISP): https://www.qualcomm.com/news/onq/2023/07/meet-the-snapdragon-8-gen-2-super-cameras
- Chips&Media y Visionary.ai, ISP completo basado en IA, 5 de enero de 2026: https://www.edge-ai-vision.com/2026/01/chipsmedia-and-visionary-ai-unveil-the-worlds-first-ai-based-full-image-signal-processor-redefining-the-future-of-image-quality/ ; entrevista de PetaPixel: https://petapixel.com/2026/01/18/a-look-at-the-worlds-first-full-ai-based-image-signal-processor
- Comité JPEG, nota de prensa de la 103.ª reunión (ganancias, complejidad e implementaciones en smartphones de JPEG AI): https://jpeg.org/items/20240507_press.html
- G. J. Sullivan, introducción a ISO/IEC JTC 1/SC 29, abril de 2026 (situación de JPEG AI; cifras de ECM y NNVC): https://www.mpeg.org/introduction-to-iso-iec-jtc-1-sc-29-coding-of-audio-picture-multimedia-and-hypermedia-information-the-world-of-jpeg-and-mpeg/
- Grupo de trabajo IEEE 1857.11 (codificación de imagen basada en redes neuronales): https://sagroups.ieee.org/fvc
- Chrome 155 y JPEG XL, 6 de octubre de 2026: https://xenospectrum.com/en/chrome-155-jpeg-xl-rust/

Llamadas y reuniones

- Microsoft, «Satin: Microsoft's latest AI-powered audio codec for real-time communications», 17 de febrero de 2021: https://techcommunity.microsoft.com/t5/microsoft-teams-blog/satin-microsoft-s-latest-ai-powered-audio-codec-for-real-time/ba-p/2141382
- Microsoft, «Announcing the Open-Source Release of ML Video Codec (MLVC)», 24 de julio de 2026: https://techcommunity.microsoft.com/blog/linuxandopensourceblog/announcing-the-open-source-release-of-ml-video-codec-mlvc/4539875 ; código: https://github.com/microsoft/mlvc
- Streaming Learning Center, «Microsoft MLVC: Interoperability, Power Realities, and the Neural Compression Patent Landscape»: https://streaminglearningcenter.com/articles/microsoft_mlvc.html ; análisis de iReplay: https://ireplay.tv/blog/mlvc-microsoft-neural-video-codec-open-source-streaming/

- Notas de la versión Opus 1.5: https://opus-codec.org/demo/opus-1.5 ; Opus 1.6: https://opus-codec.org/demo/opus-1.6
- Meta Engineering, «MLow: Meta's low bitrate audio codec», 13 de junio de 2024: https://engineering.fb.com/2024/06/13/web/mlow-metas-low-bitrate-audio-codec/
- Microsoft, PC Copilot+ (NPU de 40 TOPS o más): https://www.microsoft.com/en-us/windows/copilot-plus-pcs ; Apple, anuncio del M4 (Neural Engine de 38 TOPS): https://www.apple.com/newsroom/2024/05/apple-introduces-m4-chip/

IA generativa

- Kyutai, ficha del modelo Mimi (12,5 Hz, 1,1 kbps): https://huggingface.co/kyutai/mimi
- Lightricks, «LTX-Video: Realtime Video Latent Diffusion» (VAE de vídeo 1:192): https://arxiv.org/abs/2501.00103
- Nota de prensa de la 153.ª reunión de MPEG (Video Coding for Machines en fase DIS): https://www.mpeg.org/wp-content/uploads/2026/03/press_release_153.pdf

Videojuegos y gráficos

- NVIDIA, adopción de DLSS 4: lanzamiento y Computex 2025 (75 y más de 125): https://www.nvidia.com/en-us/geforce/news/125-dlss-4-multi-frame-gen-games-more-announced-computex-2025/ ; Gamescom 2025 (más de 175): https://www.nvidia.com/en-us/geforce/news/gamescom-2025-dlss-4-full-ray-tracing-game-announcements/ ; CES 2026 (más de 250, DLSS 4.5): https://www.nvidia.com/en-us/geforce/news/dlss-4-rtx-path-tracing-game-announcements-ces-2026/
- Ubisoft La Forge, «Shipping Neural Texture Compression in Assassin's Creed Mirage», febrero de 2026: https://www.ubisoft.com/en-us/studio/laforge/news/415slmB3ZGzv8d2LhO9QT/shipping-neural-texture-compression-in-assassins-creed-mirage
- TweakTown sobre la demo de texturas neuronales de NVIDIA en el GTC 2026 (de 6,5 GB a 970 MB): https://www.tweaktown.com/news/110886/nvidias-neural-texture-compression-cuts-vram-usage-from-6-5gb-down-to-970mb/index.html
- Blog de Microsoft DirectX, Shader Model 6.9 en versión comercial, 26 de febrero de 2026: https://devblogs.microsoft.com/directx/shader-model-6-9-retail-and-more/ ; sobre los vectores cooperativos: https://devblogs.microsoft.com/directx/shader-model-6-9-and-the-future-of-cooperative-vector/ ; planes para la GDC 2026: https://windowsreport.com/microsoft-showcases-neural-rendering-and-dxr-2-0-plans-for-directx-at-gdc-2026/

Enlaces especializados

- Global Trade Alert sobre el Anuncio n.º 49 de 2026 de China: https://globaltradealert.org/intervention/161396-china-government-revises-export-controls-on-drone-related-items ; DroneXL: https://dronexl.co/2026/10/19/china-drone-export-controls-motors-props-november-18/ ; Army Recognition (umbrales de las plataformas terrestres no tripuladas): https://www.armyrecognition.com/news/army-news/2026/china-tightens-drone-export-controls-raising-supply-chain-risks-for-the-us-europe-and-ukraine
- DroneLife, la FCC añade los drones y componentes de fabricación extranjera a la Covered List, 22 de diciembre de 2025: https://dronelife.com/2025/12/22/fcc-adds-foreign-made-drones-and-components-to-covered-list-citing-national-security-risks/

Silicio

- M. Tikekar et al., «A 249-Mpixel/s HEVC Video-Decoder Chip for 4K Ultra-HD Applications», IEEE JSSC 2014: https://dspace.mit.edu/handle/1721.1/93876 ; página del chip en el MIT (78 mW a 4K30): https://eecsg.mit.edu/chips/hevc-video-decoder-for-4k-ultra-hd-applications/
- S. Zhang et al., NVCA, DATE 2024: https://arxiv.org/abs/2312.10716
- Electronic Design, «Why Traditional VVC Will Outlive the Neural-Codec Hype on Silicon», 2026: https://www.electronicdesign.com/home/article/55414955/why-traditional-vvc-will-outlive-the-neural-codec-hype-on-silicon
- Z. Jia et al., «Towards Practical Real-Time Neural Video Compression» (DCVC-RT), CVPR 2025: https://openaccess.thecvf.com/content/CVPR2025/html/Jia_Towards_Practical_Real-Time_Neural_Video_Compression_CVPR_2025_paper.html ; material complementario (complejidad por módulo): https://openaccess.thecvf.com/content/CVPR2025/supplemental/Jia_Towards_Practical_Real-Time_CVPR_2025_supplemental.pdf
- NNVC y la biblioteca de inferencia entera SADL: https://arxiv.org/abs/2309.05846
- Tencent Canghai V2, 27 de mayo de 2026 (Shanghai Securities News vía 10jqka): https://stock.10jqka.com.cn/20260527/c677016020.shtml ; resumen de funciones: https://www.besthub.dev/articles/tencent-s-canghai-v2-chip-enters-mass-production-what-the-new-capabilities-mean-for-video-encoding-bcfc18dbfa29
- IP de decodificador AV2 de Chips&Media, 12 de mayo de 2026: https://www.design-reuse.com/news/202530501-chips-media-completes-development-of-next-gen-av2-hw-decoder-ip/ ; licencia de AV2, 17 de junio de 2026: https://www.edge-ai-vision.com/?p=58270
- Allegro DVT, IP de decodificador AV2, 16 de junio de 2026: https://www.edge-ai-vision.com/2026/06/allegro-dvts-pulsar-decoder-ip-adds-support-for-av2-video-codec/
- VeriSilicon, VC9800D con AV2, 10 de junio de 2026: https://www.businesswirechina.com/en/news/63176.html

Estándares y patentes

- Nota de prensa de la 155.ª reunión de MPEG (convocatoria conjunta de propuestas más allá de VVC), 5 de agosto de 2026: https://www.mpeg.org/wp-content/uploads/2026/08/ag3w254SN26474Press-Release-MPEG155.pdf
- Nota de prensa de la 152.ª reunión de MPEG (resultados de la convocatoria de evidencias; ediciones de VVC y VSEI), 30 de octubre de 2025: https://www.mpeg.org/wp-content/uploads/2025/11/Press-Release-MPEG152.pdf ; columna MPEG de ACM SIGMM: https://records.sigmm.org/2025/12/09/mpeg-column-152nd-mpeg-meeting/

- Streaming Media, «AI Helps InterDigital Reach Beyond VVC in Race to Develop Next-Gen Codec»: https://www.streamingmedia.com/Articles/News/Online-Video-News/AI-Helps-InterDigital-Reach-Beyond-VVC-in-Race-to-Develop-Next-Gen-Codec-171971.aspx
- GreyB, panorama de patentes más allá de VVC (H.267): https://greyb.com/blog/exclusive-video-codec-landscape/ ; sobre las licencias de AV2 y la venta de la cartera de Alibaba: https://greyb.com/blog/av2-licensing/
- MLex, «Fragmented codec licensing exposes need for clearer rules for Chinese tech majors», 27 de marzo de 2026: https://www.mlex.com/mlex/articles/2458402/fragmented-codec-licensing-exposes-need-for-clearer-rules-for-chinese-tech-majors
- Access Advance, tarifas de MCBA y VVC Advance a 1 de julio de 2026: https://accessadvance.com/wp-content/uploads/2025/07/MCBA-Program-Overview-September-2026.pdf

Dinero y empresas

- InterDigital adquiere Deep Render, 30 de octubre de 2025: https://www.advanced-television.com/2025/10/30/interdigital-acquires-deep-render
- TechCrunch, «Apple acquired a startup using AI to compress videos», 27 de marzo de 2023: https://techcrunch.com/2023/03/27/apple-acquired-a-startup-using-ai-to-compress-videos
- Neowin, «Sony acquires AI video startup iSIZE», noviembre de 2023: https://www.neowin.net/news/sony-acquires-ai-video-startup-isize-to-enhance-playstation-streaming/

COMPRESIÓN APRENDIDA · REPRODUCCIÓN × SILICIO

Codificación neuronal en 2026

Se despliega en torno al códec, no en su lugar

Cada cifra clave de este informe indica su fuente y su fecha. Cuando el autor deriva una cifra, se marca como estimación; cuando solo la comunica quien vende el producto, se marca como dato del proveedor.

Tan Shuai

Tecnólogo y fundador · Investigador y desarrollador de IA

tanshuai.com · i@tanshuai.com

EDICIÓN PÚBLICA · Datos a 12 de noviembre de 2026