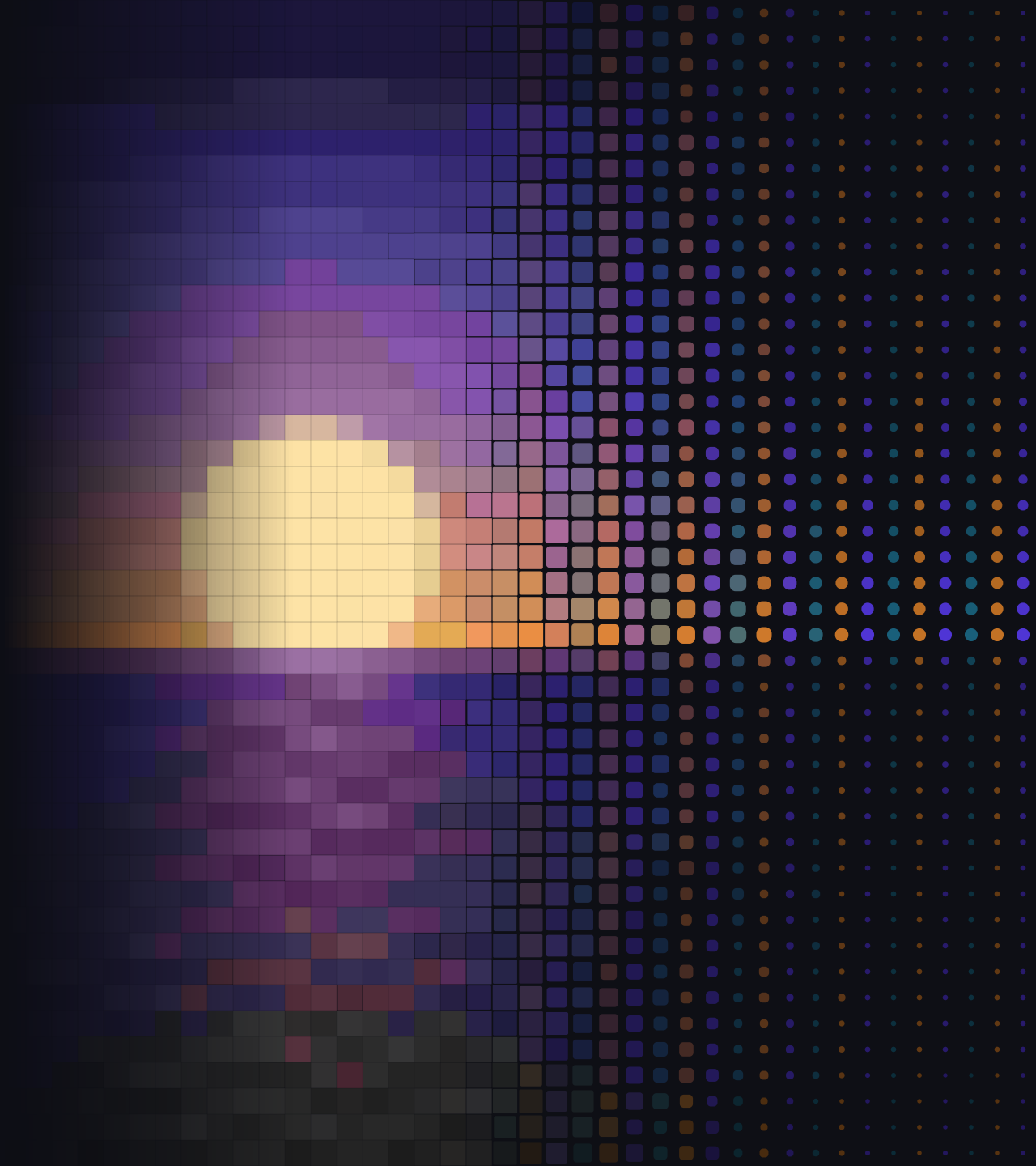
Nota di ricerca · novembre 2026

La codifica neurale nel 2026

Attorno al codec, non al suo posto

Riproduzione · Fotografia · Chiamate · IA generativa · Videogiochi · Silicio

Tan Shuai
tanshuai.com · i@tanshuai.com



IL VERDETTO

Già diffusa, ma attorno al codec

01 Le reti neurali sono già diffuse a entrambe le estremità dei codec standard

ISP delle fotocamere, downscaling lato codifica e super-risoluzione in riproduzione. Il bitstream in mezzo è ancora AV1, HEVC, AVC o VVC, decodificato da hardware a funzione fissa.

02 I codec neurali end-to-end si usano solo in circuiti chiusi

Chiamate, tokenizer dell'IA generativa e un gioco AAA: un'unica azienda controlla entrambe le estremità e può ripiegare su un codec standard.

03 I bitstream neurali sugli schermi di largo consumo aspettano H.267

La versione 1 è prevista per la fine del 2029; i primi chip verso il 2031 e l'uso di massa a metà degli anni 2030 stima.

≈75%

del traffico dati mobile è video (fine 2025)

1,26 miliardi

di smartphone consegnati nel 2025, tutti con decodifica video in hardware a funzione fissa

≈12×

l'energia per pixel di un decoder HEVC hardware, per l'unico progetto pubblicato di chip decoder neurale

Tre modi in cui una rete neurale entra in un codec



> Attorno al codec non serve il permesso di nessuno. Dentro uno standard serve uno standard. Al posto del codec serve il controllo di entrambe le estremità.

Dove si usa la codifica neurale nel 2026

	Attorno a un codec standard pre-elaborazione, miglioramento, upscaling	Codec neurale end-to-end encoder e decoder appresi	Dentro uno standard strumenti neurali in bitstream o API standard
01 **Riproduzione e streaming**	SU LARGA SCALA Downscaler di Netflix; super-risoluzione di YouTube e Meta; upscaler IA dei televisori	NON IN USO Nessun servizio di streaming invia un formato video neurale; televisori e telefoni decodificano codec standard in hardware	IN STANDARDIZZAZIONE La versione 1 di H.267 è prevista per la fine del 2029; tra i candidati ci sono strumenti neurali
02 **Fotografia e imaging**	SU LARGA SCALA ISP con IA negli smartphone di punta; un ISP interamente neurale presentato come IP al CES 2026	DEMO, PROGETTI PILOTA JPEG AI è uno standard ISO: demo su telefono, nessun prodotto	NON IN USO Il formato che guadagna terreno è classico: JPEG XL attivo di default in Chrome 155
03 **Chiamate e riunioni**	SU LARGA SCALA Super-risoluzione lato ricevente e soppressione neurale del rumore	SU LARGA SCALA Audio neurale nelle chiamate Teams tra due partecipanti dal 2021; video MLVC in distribuzione su Teams	IN USO, LIMITATO Opus 1.5 aggiunge strumenti di ML opzionali e retrocompatibili
04 **IA generativa**	—	SU LARGA SCALA Ogni modello video e vocale usa un tokenizer appreso, ad es. un VAE video 1:192	—
05 **Videogiochi e grafica**	SU LARGA SCALA Upscaling neurale e generazione di fotogrammi: il solo DLSS 4 è in oltre 250 giochi e app	IN USO, LIMITATO Texture neurali in un gioco AAA pubblicato; l'SDK di NVIDIA è in beta	IN STANDARDIZZAZIONE DirectX sta aggiungendo operazioni matriciali per gli shader (anteprima)
06 **Collegamenti specializzati: satelliti, robot**	DEMO, PROGETTI PILOTA Restauro lato server di flussi a basso bitrate	DEMO, PROGETTI PILOTA Codec neurali resistenti alle perdite in ricerca e progetti pilota	—

Le righe sono ordinate in base a quante persone raggiunge ciascun ambito.

Neurale a entrambe le estremità, standard nel mezzo

- **All’encoder**

  Il downscaler neurale di Netflix è in uso dal 2022 senza modifiche su alcun dispositivo; nei test di Netflix lo ha preferito circa il 77% degli spettatori **dato del produttore**.

- **Sullo schermo**

  Super Resolution di YouTube sui televisori da ottobre 2025; Meta usa la super-risoluzione sui propri server e sui telefoni.

- **Nel televisore**

  Il processore NQ8 AI Gen3 di Samsung dichiara 768 reti neurali per l’upscaling.

IL BITSTREAM RESTA STANDARD

30%

delle visioni su Netflix in AV1 (dicembre 2025)

90%+

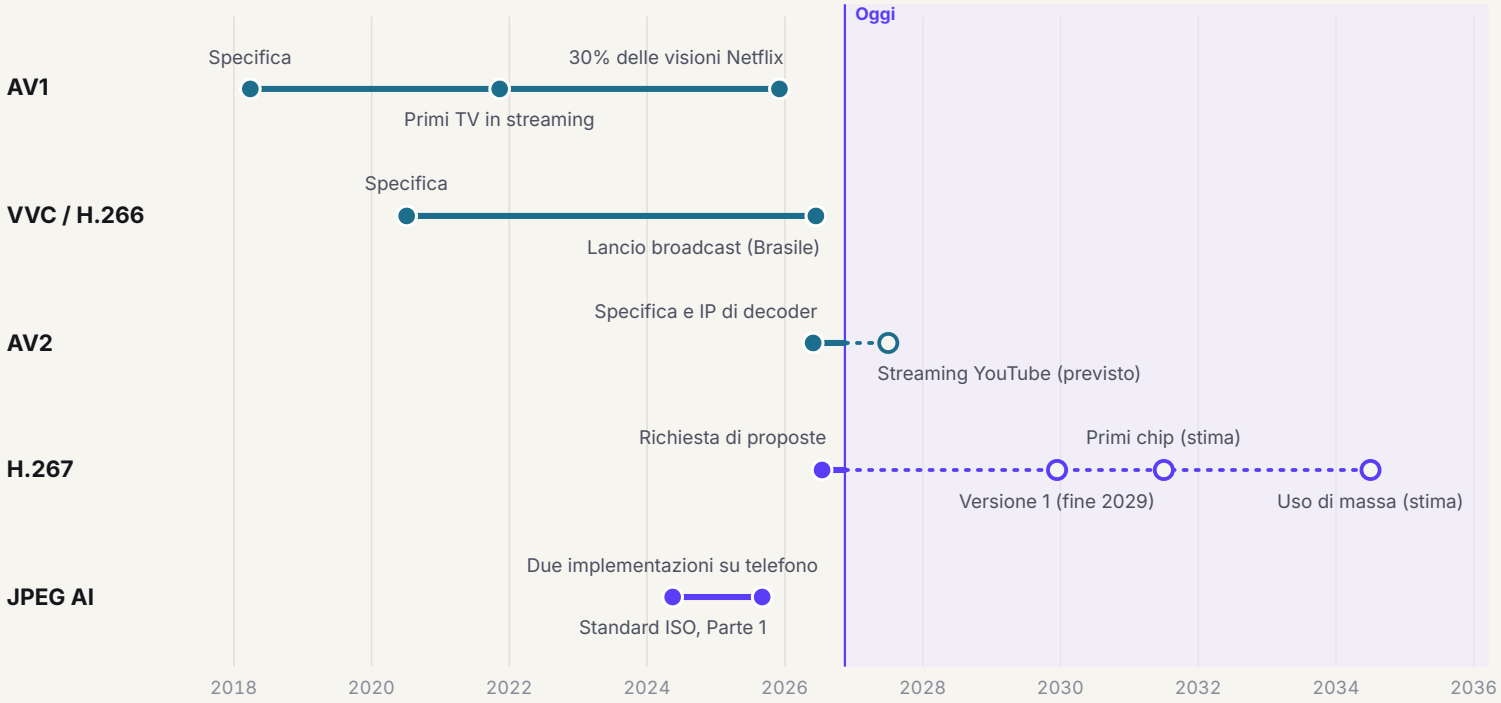
del catalogo di YouTube, per tempo di visione, disponibile in AV1 (marzo 2026)

3 settimane

da AV2 versione 1.0.0 (28 maggio 2026) ai clienti per IP di decoder di tre fornitori

TEMPI

Dallo standard finalizzato agli schermi di largo consumo: buona parte di un decennio



AV1 ha impiegato quasi otto anni dalla specifica (2018) al 30% delle visioni su Netflix (dicembre 2025). I segni vuoti indicano obiettivi o mie stime.

H.267 decide se gli strumenti neurali arriveranno su ogni schermo

Ottobre 2025
Richiesta di evidenze: cinque risposte da 16 aziende e laboratori; guadagni neurali più netti sull'8K HDR

Luglio 2026
Richiesta di proposte emessa a Ginevra

Gennaio 2027
Valutazione delle risposte a Brisbane

Seconda metà del 2027
Primo modello di test

Fine 2029
Versione 1 dello standard

≈28%
risparmiato rispetto a VVC dal filone classico ECM

6–14%
risparmiato dai filtri in-loop neurali con 10–22+ mila MAC per pixel

SEI
i flussi VVC possono già segnalare ai decoder un post-filtro neurale da applicare

La mia lettura: un nucleo classico con pochi strumenti neurali, dove i bit risparmiati giustificano il costo del decoder analisi **.**

Neurale dentro la fotocamera, classico sul web

- **Fotocamera**

  Il Cognitive ISP di Qualcomm esegue la segmentazione in tempo reale negli smartphone dal 2023; al CES 2026 è stato presentato un ISP completo con IA che gira su una NPU.

- **JPEG AI**

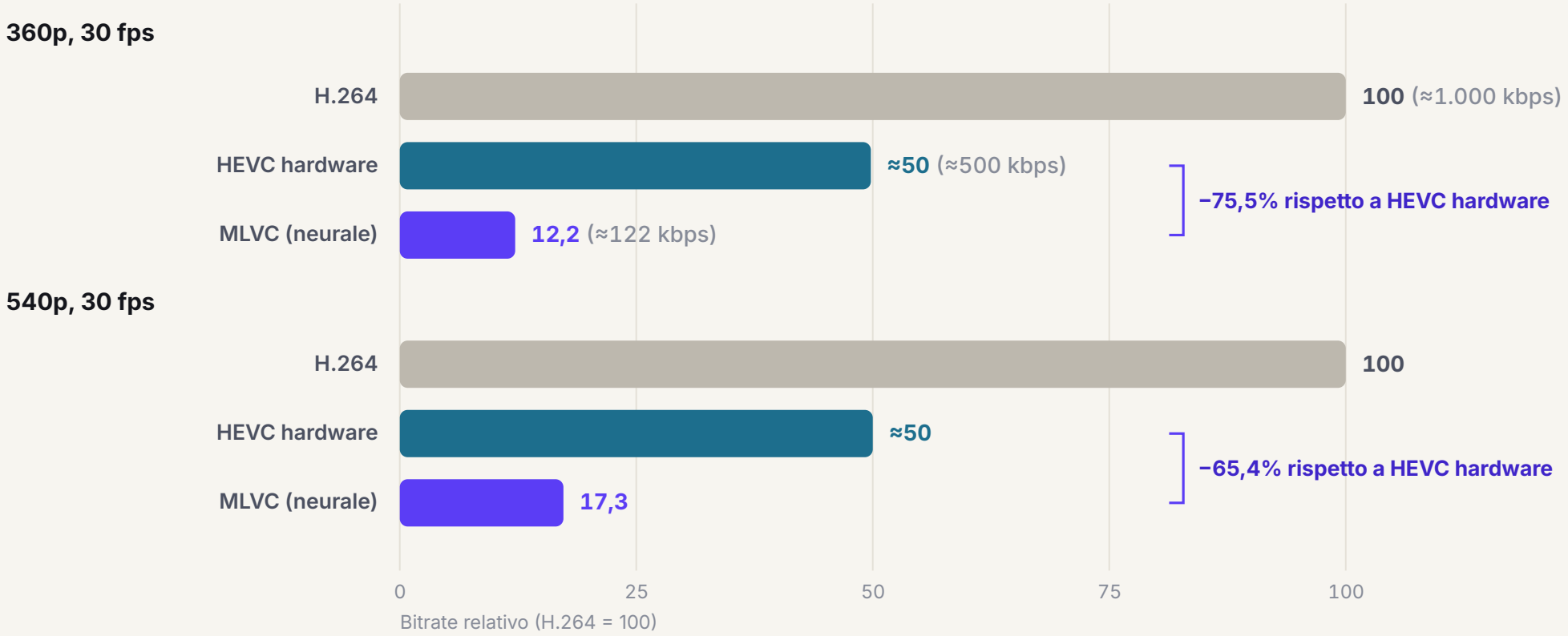
  Pubblicato nel 2025: un guadagno del 12,5–27,9% su VVC intra, con 7–216 mila MAC per pixel. Non ho trovato alcun prodotto che lo usi di default.

- **Web**

  Chrome 155 ha attivato JPEG XL di default il 6 ottobre 2026, con un decoder memory-safe scritto in Rust.

> **I formati che devono aprirsi ovunque restano classici; la codifica neurale vince dentro la pipeline di una singola azienda.**

È nelle chiamate che si usano i codec neurali end-to-end



Satin è in uso in tutte le chiamate Teams e Skype tra due partecipanti dal 2021. MLVC, open source con licenza MIT dal 24 luglio 2026, è in distribuzione nelle chiamate Teams peer-to-peer fino a 540p30, con al massimo metà della NPU **dato del produttore**.

Non tutte le piattaforme hanno scelto il neurale, e le ragioni sono istruttive

DENTRO UNO STANDARD, OPZIONALE

Opus 1.5 e 1.6

Miglioramento neurale, mascheramento della perdita di pacchetti, ridondanza profonda ed estensione di banda, tutti disattivati di default e pienamente compatibili con lo standard del 2012.

CLASSICO, PER SCELTA

Meta MLow (2024)

Oltre il 20% delle chiamate di Meta avviene su telefoni ARMv7, e decine di milioni di chiamate WhatsApp al giorno su telefoni con più di dieci anni. MLow offre circa il doppio della qualità di Opus a 6 kbps con il 10% di calcolo in meno.

Decide il budget di complessità della base installata. I codec neurali arrivano prima dove i dispositivi sono nuovi e un solo fornitore controlla entrambe le estremità.

Il più grande impiego che nessuno chiama codec

- **Codec appresi ovunque**

  Ogni modello vocale e video legge e scrive i token o le rappresentazioni latenti di un codec appreso.

- **Non formati di distribuzione**

  Ogni modello ha il proprio spazio latente: nessun controllo del bitrate, nessun test di conformità, nessun secondo decoder.

- **Una voce di costo**

  Per chi eroga video generativo, una rappresentazione latente più compressa riduce il costo di generazione più che proporzionalmente analisi.

DUE ESEMPI

12,5 Hz · 1,1 kbps

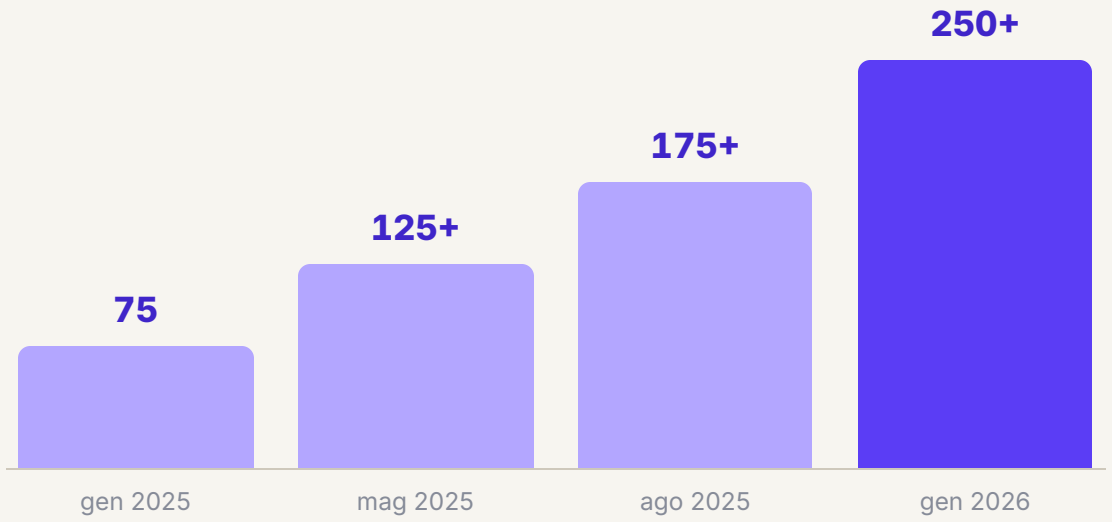
i token vocali Mimi di Kyutai, sotto qualsiasi codec telefonico

1:192

il VAE video di LTX-Video: un token latente per 32×32 pixel su 8 fotogrammi

La ricostruzione neurale è diffusa; gli asset neurali sono agli inizi

DLSS 4 Multi Frame Generation: giochi e app



Compressione neurale delle texture

Scena demo NVIDIA, GTC 2026
Memoria texture 6,5 GB → 0,97 GB

85% di memoria in meno

Assassin's Creed Mirage, pubblicato
Solo asset selezionati; rispetto ai formati a blocchi

≈30% di compressione in più

Misure diverse: non direttamente confrontabili

DirectX Linear Algebra è entrato in anteprima pubblica ad aprile 2026: un unico percorso di codice per la decodifica neurale negli shader, sulla GPU di qualsiasi produttore.

Satelliti, droni e robot: l’uso più adatto, un mercato piccolo

- **Perché si adatta**

  Un unico proprietario controlla entrambe le estremità, la banda è scarsa e il contenuto è abbastanza circoscritto da poterci addestrare una rete.

- **Cina**

  L’Annuncio n. 49 del 2026, in vigore dal 18 novembre, rivede i controlli all’esportazione dei droni; una voce usa come soglia la compressione semantica delle immagini.

- **Stati Uniti**

  Il 22 dicembre 2025 la FCC ha aggiunto alla sua Covered List i droni prodotti all’estero e i loro componenti critici.

> **Una nicchia solida per uno specialista capace di gestire la conformità all’export, non un mercato di massa.**

Un ordine di grandezza in più di energia per pixel

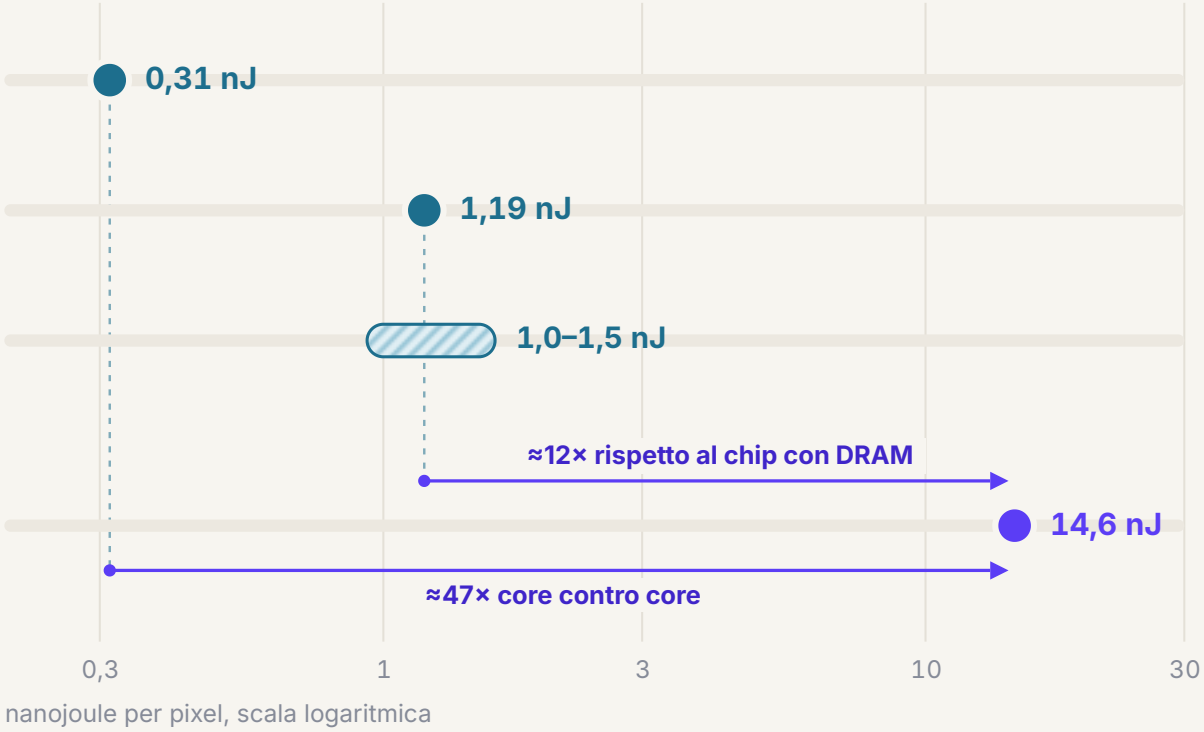
Chip decoder HEVC, core
40 nm, 4K30, misurato

Chip decoder HEVC, con DRAM
40 nm, 4K30, core misurato più DRAM simulata

Decoder VVC
28 nm, 4K60, proiezione

Decoder video neurale (NVCA)
28 nm, 1080p25, solo sintesi; pesi e DRAM esclusi

A 1080p30 sono circa 74 mW contro 908 mW.



A 1080p30: circa 74 mW per il chip HEVC con DRAM contro all'incirca 900 mW per NVCA. Nodo più vecchio per il chip HEVC, pesi e DRAM esclusi per NVCA: il confronto favorisce il progetto neurale.

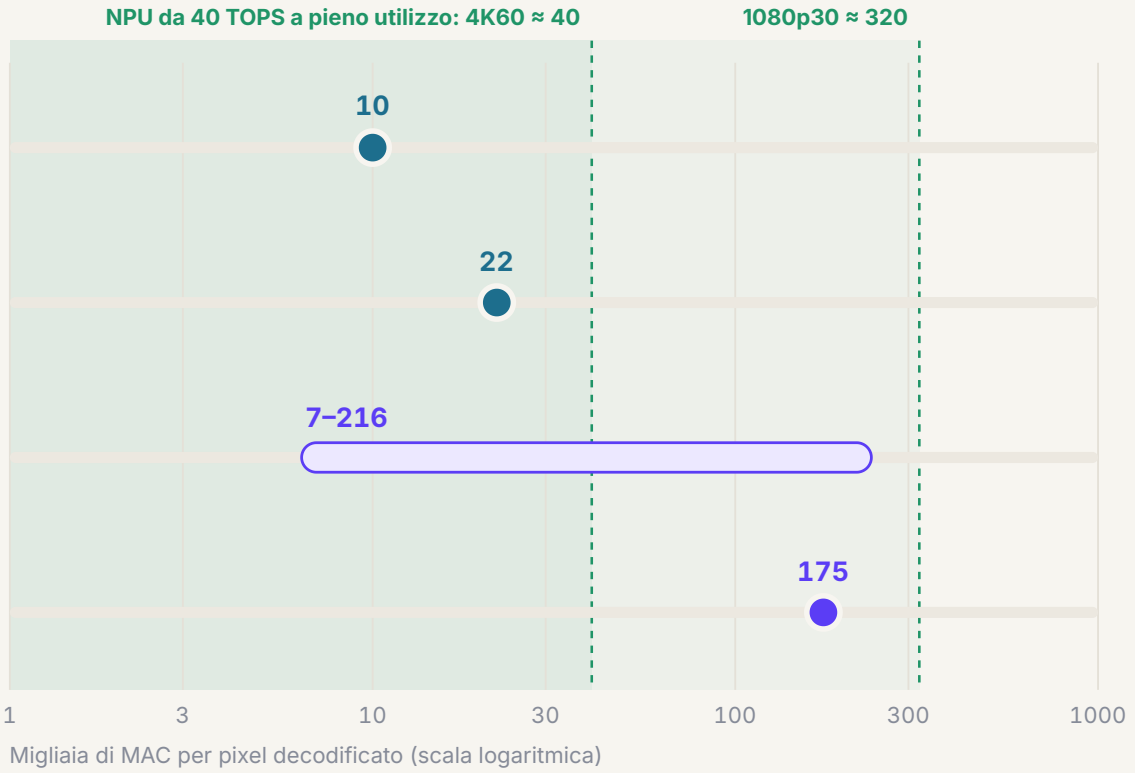
Che cosa chiede un decoder neurale a una NPU

Filtro in-loop neurale, complessità molto bassa
Esplorazione JVET · ≈6% di bit risparmiati
1,2 TOPS a 1080p30 · 10 a 4K60

Filtro in-loop neurale, bassa complessità
Esplorazione JVET · ≈8%
2,7 TOPS a 1080p30 · 22 a 4K60

Decoder JPEG AI, livelli da base ad alto
12,5–27,9% rispetto a VVC intra

DCVC-RT, un intero decoder video neurale
21% rispetto al riferimento VVC, condizioni dell'articolo
22 TOPS a 1080p30 · 174 a 4K60



DCVC-RT richiede circa 175 mila MAC per pixel: circa 22 TOPS a 1080p30 e 174 a 4K60, a pieno utilizzo **stima**. Una NPU da 40 TOPS a pieno utilizzo può permettersene circa 40 mila a 4K60.

Traffico di memoria ed esattezza al bit sono i vincoli più duri

DOMINA LA DRAM

La memoria prima dei MAC

Nel chip HEVC del MIT la DRAM vale circa tre quarti dell'energia del sistema: 0,88 su 1,19 nJ per pixel. I decoder neurali portano grandi mappe di feature da un fotogramma all'altro, e per gli autori di DCVC-RT a limitare la velocità era l'I/O di memoria, non l'aritmetica.

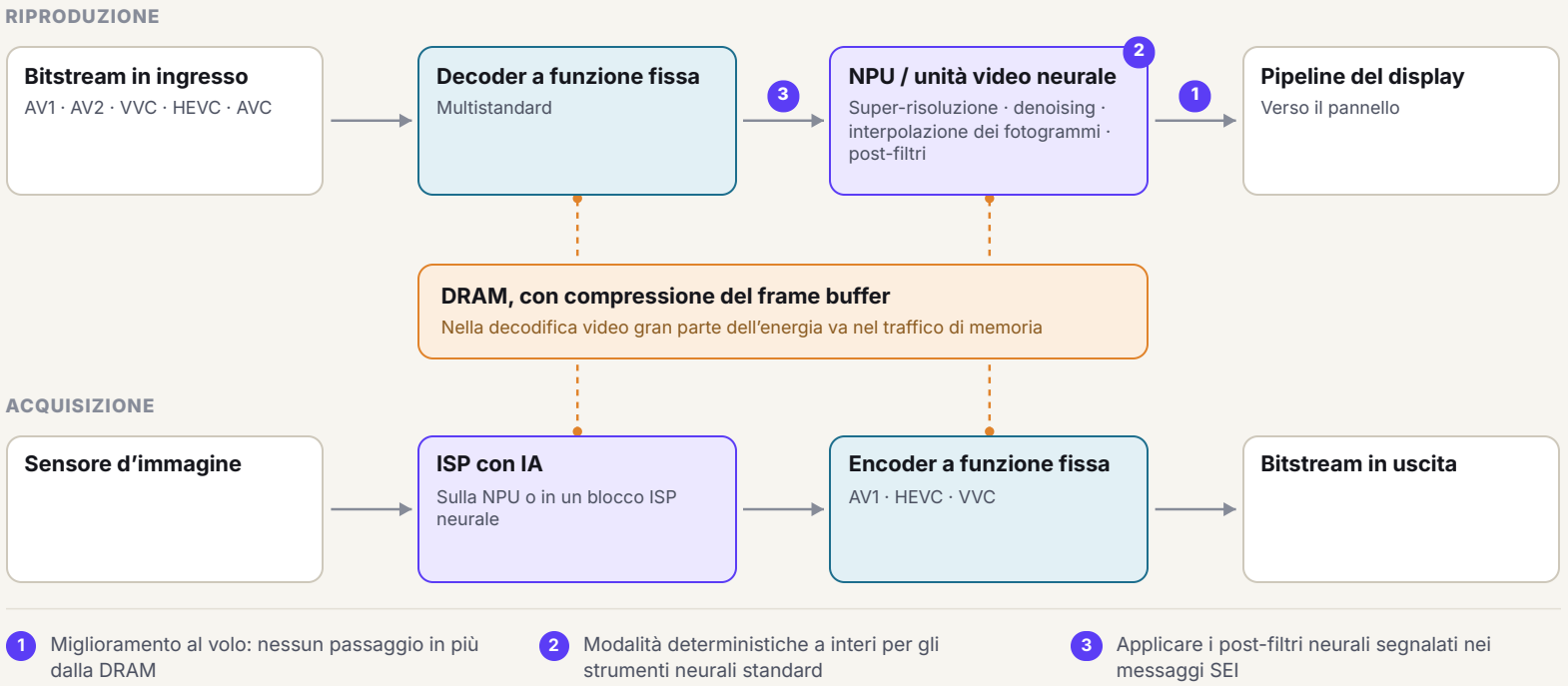
ESATTEZZA

Ogni decoder deve calcolare gli stessi bit

La decodifica entropica fallisce se due dispositivi calcolano una probabilità in modo diverso. Le soluzioni: modelli a interi (DCVC-RT), un hyperprior a scala condivisa (MLVC), inferenza su interi a 16 bit (SADL di JVET), punti di conformità (JPEG AI).

Per un team di chip è di nuovo la disciplina delle trasformate intere di HEVC: lo standard deve definire l'aritmetica in modo esatto.

La ripartizione che arriva nei prodotti: un codec a funzione fissa più una NPU



> Canghai di Tencent: oltre 100.000 chip di prima generazione installati; la seconda generazione aggiunge H.266 e la super-risoluzione sul chip. Chips&Media e Allegro DVT vendono IP di codec e IP video neurali fianco a fianco.

Cosa costruirei, per ruolo

Team di SoC e chip per TV

- Decoder a funzione fissa per ogni standard da riprodurre, AV2 subito
- Una NPU dimensionata per il miglioramento alla risoluzione del display, senza un passaggio in più in DRAM
- SEI per post-filtri neurali e modalità deterministiche a interi
- H.267 come avvio di progetto nel 2029 o più tardi

Fornitori di IP

- Decoder AV2 e VVC subito
- Acceleratori di post-filtri e compressione del frame buffer
- Strumenti di conformità per l'elaborazione neurale

Piattaforme video

- La pre-elaborazione neurale all'encoder ripaga per prima
- Super-risoluzione lato client dove i dispositivi lo permettono
- Codec neurali end-to-end solo in circuiti chiusi sotto il proprio controllo

Dove gli strumenti neurali diventano oggetto di licenza

Standard	Contenuto neurale	Stato, novembre 2026
H.267 (nome provvisorio)	Aperto: strumenti classici (ECM) e neurali sono entrambi candidati	Risposte valutate a gennaio 2027; versione 1 a fine 2029
VVC e HEVC, con VSEI	Post-filtro neurale descritto in SEI; facoltativo per i decoder	In vigore; esteso a ottobre 2025
AV2	Classico, esente da royalty	Versione 1.0.0 il 28 maggio 2026; stream web di YouTube previsti per inizio 2027
JPEG AI	Codec d'immagine appreso end-to-end	Parte 1 pubblicata nel 2025; nessun supporto di default nei prodotti trovato
IEEE 1857.11	Codifica neurale di immagini progettata per NPU	Pubblicato a dicembre 2024
Video Coding for Machines	Codifica per l'analisi automatica anziché per la visione	Bozza di standard internazionale a inizio 2026
Estensioni ML di Opus	Opzionali e retrocompatibili	In libopus 1.5 e 1.6

645

famiglie di brevetti post-VVC detenute da ByteDance (GreyB, 2025)

0,50 $ · 1,00–1,50 $

tariffa del pool VVC per telefono e per TV 4K, Region 1

0 $

di royalty per AV1 e AV2 dai membri di AOMedia

CHI GUADAGNA

Il denaro segue gli standard e le piattaforme

GUADAGNANO OGGI

Piattaforme

Si tengono i risparmi

Downscaler neurale di Netflix; super-risoluzione di YouTube e Meta; Microsoft ha reso open source MLVC con licenza MIT

GUADAGNANO OGGI

Produttori di dispositivi e chip

Vendono differenziazione

ISP delle fotocamere; processori di upscaling per TV; i chip codec Canghai sviluppati da Tencent

GUADAGNANO OGGI

Fornitori di IP

Licenziano ogni nuovo standard

Clienti per decoder AV2 presso Chips&Media, Allegro DVT e VeriSilicon entro tre settimane dalla specifica

GUADAGNANO OGGI

Produttori di GPU

Creano dipendenza dalla piattaforma

DLSS 4 in oltre 250 giochi e app; SDK per texture neurali in beta

PER DISPOSITIVO

Titolari di brevetti e pool

Incassano per dispositivo

VVC: 0,50 $ per telefono, 1,00–1,50 $ per TV 4K; strumenti H.267 dopo il 2029

USCITA PER ACQUISIZIONE

Startup di codec neurali

Acquisite

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

> Le tre startup di codec neurali più note, WaveOne, iSIZE e Deep Render, sono state acquistate da Apple, Sony e InterDigital.

DOVE PUÒ INSERIRSI UN'AZIENDA

Sei opportunità, e il rischio di ciascuna

Opportunità	Chi compra	Perché ora	Rischio principale
Pre-elaborazione neurale all'encoder	Piattaforme di streaming e di video degli utenti, broadcaster	Nessuna modifica ai dispositivi; collaudata da Netflix	Le grandi piattaforme se la costruiscono da sole
IP e taratura per il miglioramento lato display	Produttori di chip per TV, set-top box e telefoni	Le NPU sono già nei dispositivi	Commoditizzazione; team interni
Codec neurali a circuito chiuso	Fornitori di comunicazioni aziendali, industriali e di telepresenza	MLVC dimostra che funziona, e il suo codice è aperto	Occorre controllare entrambe le estremità; frammentazione delle NPU
Silicio o IP codec più NPU	Produttori di SoC	Il ciclo di AV2 sta iniziando	Operatori già affermati e forti
Valutazione, conformità e inferenza deterministica	Chiunque porti strumenti neurali nei prodotti	L'esattezza al bit è un requisito inderogabile	Mercato piccolo se non abbinato ad altro
Contributi alla standardizzazione di H.267	Licenzianti, grandi piattaforme	Le risposte si valutano a gennaio 2027	Lungo, costoso, incerto

AFFERMAZIONI RICORRENTI

Diffidare di un’affermazione finché non dichiara quattro cose

Affermazione	Che cosa dicono le prove
«I codec neurali battono già VVC»	In laboratorio. DCVC-RT risparmia il 21% rispetto all’encoder di riferimento di VVC; il suo decoder costa circa 175 mila MAC per pixel.
«I codec IA sostituiranno H.26x e AV1 entro pochi anni»	Non nella riproduzione: versione 1 di H.267 a fine 2029, chip verso il 2031 stima. Il codec che arriva ora, AV2, è classico.
«MLVC riduce la banda dell’88%»	Rispetto a H.264 a 360p. Rispetto a HEVC hardware: 75,5% a 360p e 65,4% a 540p, sul set di test della stessa Microsoft.
«Le texture neurali riducono la memoria video dell’85–96%»	Nelle scene dimostrative. Il gioco pubblicato ha guadagnato circa il 30%, solo su asset selezionati.
«JPEG AI è uno standard ISO, quindi fotocamere e browser lo adotteranno»	Nessun supporto di default trovato nei prodotti. Il web ha adottato JPEG XL, un formato classico.

Chiedere il riferimento, la metrica, la risoluzione e il costo di decodifica. Un’affermazione che non sa indicarli tutti e quattro è un risultato di ricerca, non un prodotto.

Cosa osservare, e le questioni aperte

INDICATORI

Gennaio 2027	Valutazione delle risposte per H.267: quante si basano su strumenti neurali, a quale costo per pixel
Inizio 2027	I primi stream web AV2 di YouTube; decoder AV2 nei SoC annunciati
Seconda metà del 2027	Il primo modello di test di H.267
2027	MLVC oltre le chiamate Teams peer-to-peer
2027–2028	Un primo prodotto che usa JPEG AI di default
Fine 2029	La versione 1 di H.267 avvia il ciclo del silicio

QUESTIONI APERTE

01 Esattezza su larga scala: decodifica a interi esatta al bit sulle NPU di produttori diversi, con quale perdita di compressione?

02 La memoria prima dei MAC: il traffico DRAM per pixel della post-elaborazione neurale

03 Il costo per pixel oltre il quale uno strumento neurale smette di ripagarsi in H.267

04 Quali circuiti chiusi, oltre alle chiamate, adotteranno per primi i codec neurali end-to-end

05 Come misurare la fedeltà quando un decoder genera dettagli

Sono temi che possiamo esplorare insieme, attraverso lo scambio tecnico e la collaborazione tecnica.

i@tanshuai.com · tanshuai.com