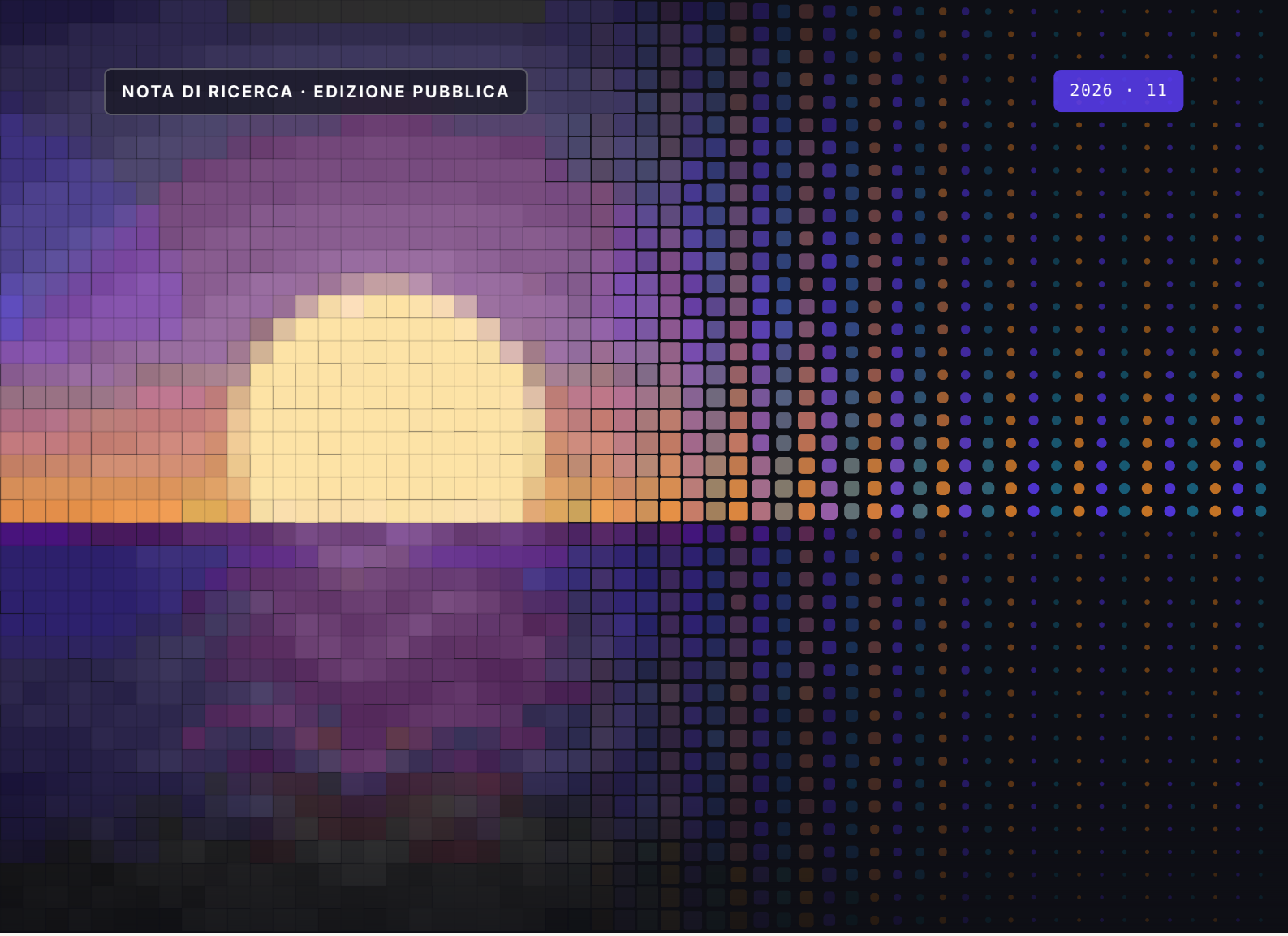


NOTA DI RICERCA · EDIZIONE PUBBLICA

2026 · 11

COMPRESSIONE APPRESA · RIPRODUZIONE × SILICIO

La codifica neurale nel 2026

Si usa attorno al codec, non al suo posto

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

IL VERDETTO

Le reti neurali lavorano già a entrambe le estremità dei codec che trasportano i video di tutto il mondo. I codec neurali end-to-end si usano solo in circuiti chiusi, e gli schermi di largo consumo aspettano H.267.

Tan Shuai

Tecnologo e fondatore · Ricercatore e sviluppatore di IA

tanshuai.com · i@tanshuai.com

Dati aggiornati al 12 novembre 2026

Importi in dollari USA

INDICE

Indice

◆ **Sintesi** — Già diffusa, ma attorno al codec — 04

01 **Ambito e metodo** — Che cosa significa qui «codifica neurale», e come l'ho esaminata — 07

02 **Riproduzione e streaming** — Riproduzione e streaming: neurale a entrambe le estremità, standard nel mezzo — 08

03 **Fotografia e immagini** — Fotografia e immagini: neurale dentro la fotocamera, classico sul web — 11

04 **Chiamate e riunioni** — Chiamate e riunioni: dove i codec neurali end-to-end si usano davvero — 12

05 **IA generativa** — IA generativa: il più grande impiego che nessuno chiama codec — 14

06 **Videogiochi e grafica** — Videogiochi e grafica: la ricostruzione neurale è diffusa, gli asset neurali sono agli inizi — 15

07 **Collegamenti specializzati** — Collegamenti specializzati: satelliti, droni e robot — 17

08 **Silicio** — La prospettiva del silicio: energia, calcolo, memoria ed esattezza — 18

09 **Standard e brevetti** — Standard e brevetti: dove gli strumenti neurali diventano oggetto di licenza — 23

10 **Chi guadagna** — Chi guadagna, e dove può inserirsi un'azienda — 24

11 **Affermazioni e prove** — Affermazioni ricorrenti, e che cosa dicono le prove — 26

12 **Indicatori** — Cosa osservare, dal 2027 al 2030 — 27

13 **Questioni aperte** — 28

◆ **L'autore** — Tan Shuai — 29

◆ **Fonti** — 30

FIGURE

Figure

FIG. 01	Dove si usa la codifica neurale nel 2026	06
FIG. 02	Tre modi in cui una rete neurale entra in un codec	07
FIG. 03	Dallo standard finalizzato agli schermi di largo consumo: buona parte di un decennio	10
FIG. 04	Nelle chiamate un codec neurale batte HEVC hardware del 65–76%	13
FIG. 05	Quando un'unica azienda controlla l'intero stack, la ricostruzione neurale si diffonde in fretta	16
FIG. 06	Energia per pixel decodificato: i decoder a funzione fissa sono avanti di un ordine di grandezza	19
FIG. 07	Uno strumento neurale costa decine di migliaia di MAC per pixel; un intero decoder neurale centinaia di migliaia	20
FIG. 08	La ripartizione che arriva nei prodotti: hardware a funzione fissa per il bitstream, un motore neurale per i pixel	21
FIG. 09	Chi guadagna con la codifica neurale nel 2026	24

Etichette nel testo

- **stima** — calcolo o deduzione dell'autore a partire da dati citati
- **dato del produttore** — pubblicato dall'azienda che vende il prodotto, non verificato in modo indipendente
- **analisi** — interpretazione delle prove da parte dell'autore

Questa è una traduzione dell'edizione inglese, che prevale in caso di differenze.

Dati aggiornati al 12 novembre 2026 · Importi in dollari USA

SINTESI

Già diffusa, ma attorno al codec

IL VERDETTO

La codifica neurale è già diffusa nel 2026, ma lavora attorno ai codec standard, non al loro posto.

≈75%
del traffico dati mobile è video, fine 2025 (Ericsson)

1,26 miliardi
di smartphone consegnati nel 2025, ognuno con decodifica video in hardware a funzione fissa

≈12×
l'energia per pixel di un chip decoder HEVC misurato, con DRAM, per l'unico progetto pubblicato di chip decoder neurale

≈22 TOPS
di calcolo sostenuto per un decoder neurale di classe DCVC-RT a 1080p30 stima

75,5%
di bitrate in meno rispetto a HEVC hardware per MLVC a 360p nelle chiamate, sul set di test della stessa Microsoft dato del produttore

Fine 2029
versione 1 di H.267, il primo standard che probabilmente porterà strumenti neurali su ogni schermo

Dove si usa

Le reti neurali lavorano ormai a entrambe le estremità dei codec che trasportano quasi tutto il video del mondo. In ingresso modellano l'immagine: processori di segnale d'immagine con IA nelle fotocamere degli smartphone e un downscaler neurale nella pipeline di codifica di Netflix. In uscita la riparano: la Super Resolution di YouTube sui televisori, la super-risoluzione di Meta e processori per TV che dichiarano centinaia di reti neurali. In mezzo, il bitstream è ancora AV1, HEVC, AVC o VVC, decodificato da hardware a funzione fissa. Nessun servizio di streaming invia un formato video neurale a televisori o telefoni; dove oggi un decoder video neurale gira su un dispositivo, gira dentro l'app di una sola azienda.

Circuiti chiusi

I codec neurali end-to-end, in cui sia l'encoder sia il decoder sono appresi, si usano solo in circuiti chiusi, dove un'unica azienda controlla entrambe le estremità e può ripiegare su un codec standard. Dal 2021 Microsoft usa un codec audio basato sull'IA in tutte le chiamate Teams e Skype tra due partecipanti, e a luglio 2026 ha reso open source MLVC, un codec video neurale ora in distribuzione nelle chiamate Teams peer-to-peer: a 360p richiede circa il 75% di bitrate in meno rispetto a HEVC hardware a parità di qualità valutata. Ogni modello di generazione vocale e video si basa su un codec appreso, dai tokenizer vocali a 12,5 Hz ai VAE video con un rapporto di compressione di 1:192. La compressione neurale delle texture è arrivata in un gioco AAA. Sono impieghi reali, e hanno una condizione in comune: nessun altro deve decodificare il flusso.

Tempi

La riproduzione di massa di bitstream neurali aspetta uno standard e un ciclo del silicio. Il primo standard che probabilmente porterà strumenti neurali, il successore di VVC che

chiamerò H.267, ha emesso la sua richiesta di proposte a luglio 2026, valuterà le risposte a gennaio 2027 e punta a una prima versione per la fine del 2029. AV1 ha impiegato quasi otto anni dalla specifica finalizzata al 30% delle visioni su Netflix. Seguendo lo stesso schema, i primi chip H.267 arriveranno verso il 2031 e l'uso di massa a metà degli anni 2030 stima. Nel frattempo, il prossimo codec ad arrivare sugli schermi è AV2, finalizzato a maggio 2026 senza alcun decoder neurale al suo interno, e tre fornitori di IP hanno annunciato clienti per il decoder AV2 entro tre settimane.

Silicio

La ragione sta nel silicio. L'unico progetto pubblicato di chip decoder video neurale richiede 14,6 nJ per pixel, circa 12 volte un chip decoder HEVC misurato, conteggiato con la sua DRAM, e circa 47 volte il suo solo core. DCVC-RT, il codec di ricerca in tempo reale di Microsoft, richiede, secondo il mio conteggio sulla sua stessa tabella per modulo, circa 175 mila moltiplicazioni-accumulo per pixel decodificato, ovvero all'incirca 22 TOPS di calcolo sostenuto per il 1080p30 stima: circa la metà dei 40 TOPS che definiscono un PC con IA, spesi per un solo flusso video. Un decoder neurale deve anche produrre risultati identici bit per bit su ogni dispositivo, altrimenti la sua decodifica entropica va in pezzi. Per questo MLVC rinuncia a parte della compressione per decodificare in modo identico su NPU diverse, e funziona a 540p, non in 4K.

Mercato

Per un'azienda, il denaro segue gli standard e le piattaforme, non i codec neurali a sé stanti. Le piattaforme si tengono i risparmi della pre- e post-elaborazione neurale. I fornitori di IP concedono in licenza decoder a funzione fissa per ogni nuovo standard, più NPU per l'imaging. I pool di brevetti incassano per dispositivo: 0,50 dollari per telefono e 1,00–1,50 dollari per TV 4K per VVC negli Stati Uniti e negli altri mercati della Region 1. Le tre startup di codec neurali più note, WaveOne, iSIZE e Deep Render, sono state acquistate da Apple, Sony e InterDigital. Le opportunità aperte sono nella pre- e post-elaborazione neurale, nei prodotti a circuito chiuso, nel silicio che abbina un codec a funzione fissa a una NPU, e nel lavoro di valutazione e di standardizzazione che sta in mezzo.

Domanda	Giudizio	Numeri chiave (fonti nei capitoli)
① **Dove si usa oggi la codifica neurale su larga scala?**	Attorno al codec: ISP delle fotocamere, downscaling lato codifica, super-risoluzione in riproduzione. End-to-end solo in circuiti chiusi: chiamate, asset di gioco, tokenizer dell'IA generativa	Video ≈75% dei dati mobili; 1,26 miliardi di telefoni e ≈200 milioni di televisori consegnati nel 2025, tutti con decodifica in hardware a funzione fissa; DLSS 4 in oltre 250 giochi e app
② **I codec neurali sostituiranno AV1, HEVC e VVC nella riproduzione? Quando?**	Non prima di uno standard e di un ciclo di dispositivi. H.267 è il primo veicolo realistico, e probabilmente mescolerà strumenti classici e neurali	Richiesta di proposte H.267 luglio 2026, valutazione gennaio 2027, versione 1 fine 2029; uso di massa a metà degli anni 2030 stima
③ **Quanto costa in silicio?**	Un ordine di grandezza in più di energia per pixel rispetto all'attuale decodifica a funzione fissa, più il traffico di memoria e un requisito di esattezza al bit	14,6 nJ/pixel (neurale) contro 1,19 (HEVC, con DRAM); ≈175 kMAC per pixel decodificato ≈ 22 TOPS a 1080p30 stima
④ **Che cosa permettono standard e brevetti?**	I post-filtri neurali possono già essere segnalati nei flussi VVC; JPEG AI è uno standard ISO; H.267 deciderà quali strumenti neurali diventeranno oggetto di licenza	ECM (classico) ≈28% su VVC; filtri in-loop neurali 6–14% a 10–22+ kMAC/pixel; ByteDance 645 famiglie di brevetti nel lavoro post-VVC
⑤ **Chi guadagna, e dove può inserirsi un'azienda?**	Piattaforme, fornitori di IP, titolari di brevetti. Le startup dedicate solo ai codec neurali sono state acquisite. Opportunità in pre/post-elaborazione, circuiti chiusi, silicio codec più NPU, valutazione	Pool VVC 0,50 dollari per telefono; IP di decoder AV2 concessa in licenza a poche settimane dalla specifica; tre acquisizioni, condizioni non rese note

Domanda	Giudizio	Numeri chiave (fonti nei capitoli)
⑥ **Di quali affermazioni diffidare?**	Guadagni dichiarati rispetto a codec vecchi, a bassa risoluzione, senza costo di decodifica, o tratti da demo anziché da prodotti in commercio	MLVC: 88% in meno di H.264 ma 65–76% in meno di HEVC hardware; texture: 85% di VRAM risparmiata in una demo, circa 30% nel gioco pubblicato

FIG. 01

Dove si usa la codifica neurale nel 2026

Le reti neurali lavorano già a entrambe le estremità dei codec più diffusi. I codec neurali end-to-end si usano solo dove un'unica azienda controlla sia l'encoder sia il decoder.

	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	—

Fonte: Documenti di aziende, enti di standardizzazione e sviluppatori citati nel rapporto; situazione al 12 novembre 2026. Le righe sono ordinate in base a quante persone raggiunge ciascun ambito.

AMBITO E METODO

01 Che cosa significa qui «codifica neurale», e come l'ho esaminata

Tre modi in cui una rete neurale entra in un codec

«Codifica neurale» indica tre cose diverse, e gran parte delle discussioni in materia nasce dal confonderle.

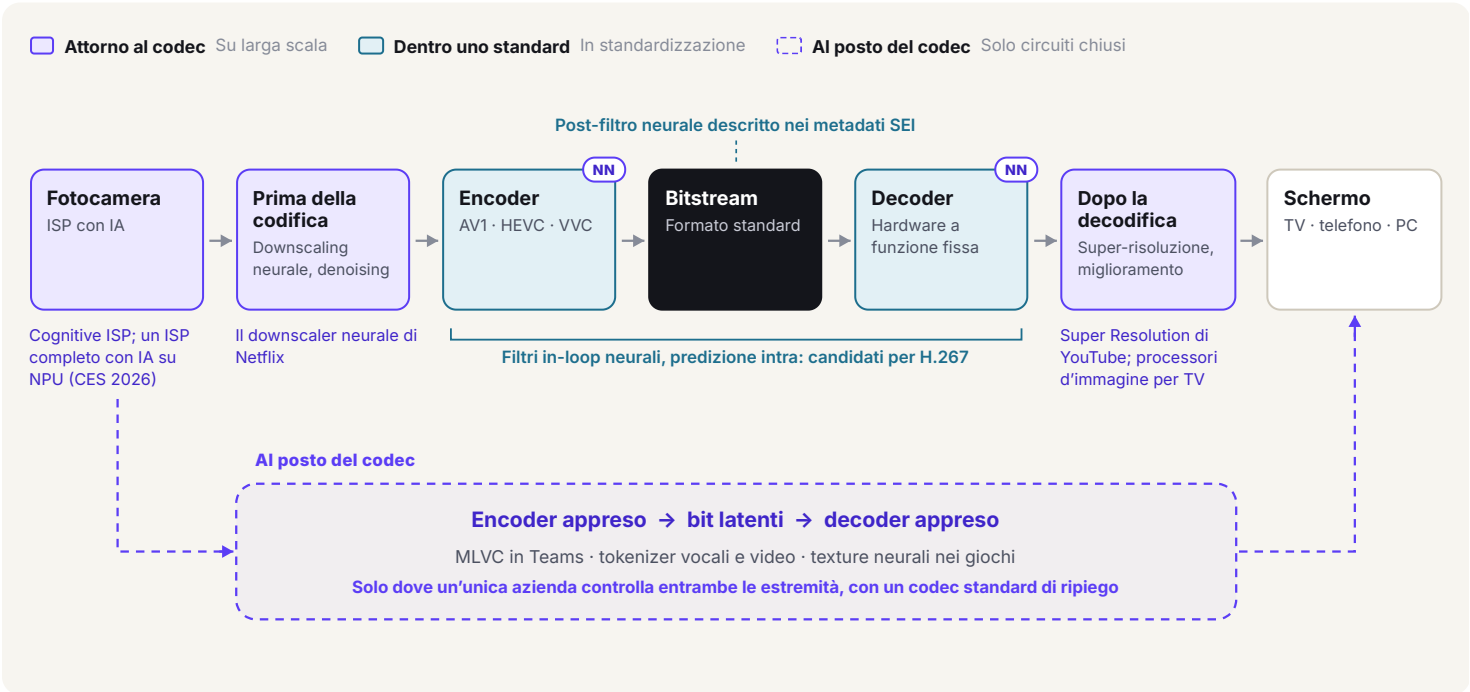
01 **Attorno al codec.** Una rete elabora i pixel prima della codifica o dopo la decodifica, e il bitstream resta standard. Esempi: un processore di segnale d'immagine (ISP) con IA in una fotocamera, il downscaler neurale di Netflix prima del suo encoder, la super-risoluzione su un televisore. Qualsiasi dispositivo può riprodurre il flusso; la rete cambia soltanto ciò che entra o ciò che esce.

02 **Dentro uno standard.** Una rete diventa uno strumento di un codec standard, per esempio un filtro in-loop o un predittore intra neurale che ogni decoder deve eseguire, e quindi deve essere specificata in modo esatto, di solito in aritmetica intera. Una forma più leggera è un post-filtro neurale che il bitstream descrive nei metadati e che i decoder possono scegliere di applicare.

03 **Al posto del codec.** Sia l'encoder sia il decoder sono appresi end-to-end. Il bitstream è un insieme di valori latenti quantizzati che solo la rete corrispondente può decodificare. Esempi: MLVC di Microsoft in Teams, i tokenizer vocali nei modelli di voce, le texture neurali in un gioco.

FIG. 02 **Tre modi in cui una rete neurale entra in un codec**

Gran parte di ciò che si usa nel 2026 lavora attorno a un codec standard. Gli strumenti neurali dentro uno standard sono in fase di specifica. I codec neurali end-to-end sostituiscono il codec solo dove un'unica azienda controlla entrambe le estremità.



Fonte: Sintesi dell'autore delle fonti di questo rapporto; situazione al 12 novembre 2026.

Il primo tipo non ha bisogno del permesso di nessuno. Il secondo ha bisogno di uno standard. Il terzo ha bisogno del controllo di entrambe le estremità. È l'ordine in cui arrivano alle persone, e spiega gran parte di ciò che segue.

Ordinati secondo il mondo reale, non secondo l'agenda della ricerca

Gli articoli di ricerca studiano soprattutto il terzo tipo, perché è lì che si trovano i record di compressione. Questo rapporto ordina gli ambiti in base a quante persone raggiungono, perché è questo che decide dove vanno il denaro e l'ingegneria: prima la riproduzione e lo streaming, poi la fotografia, le chiamate, l'IA generativa, i videogiochi e, per ultimi, i collegamenti specializzati (satelliti, droni, robot). Il video rappresenta circa il 75% di tutto il traffico dati mobile, secondo il conteggio di Ericsson alla fine del 2025. Nel 2025 l'industria ha consegnato circa 1,26 miliardi di smartphone (IDC) e all'incirca 196–209 milioni di televisori (stime di TrendForce e Omdia). Ognuno di questi dispositivi decodifica il video in hardware a funzione fissa.

Come ho lavorato

Questa è una ricerca documentale, condotta come esaminerei un chip o un piano di prodotto. Ho letto documenti di standardizzazione e resoconti di riunioni (JVET, MPEG, JPEG, AOMedia, IETF), blog di ingegneria e note di rilascio delle aziende, articoli sui chip, tariffari dei pool di brevetti, testi sui controlli all'esportazione e la stampa. Ogni numero riporta la sua data. Dove ricavo io stesso un numero, per esempio il calcolo per pixel a partire dalle tabelle di un articolo, lo dico e lo contrassegno come **stima**. Le affermazioni che non ho potuto verificare su una fonte primaria sono contrassegnate come **dato del produttore**. Gli importi sono in dollari USA. La situazione è aggiornata al 12 novembre 2026.

Confronto grandezze omogenee dove posso, e lo segnalo dove non posso. I guadagni di compressione significano poco se non sono accompagnati da quattro cose: il riferimento (guadagno rispetto a che cosa), la metrica (PSNR, VMAF o un panel di spettatori), il contenuto e la risoluzione, e quanto costa la decodifica. Gran parte del disaccordo sui codec neurali scompare una volta dichiarate queste quattro cose.

CAPITOLO 02

RIPRODUZIONE E STREAMING

02 Riproduzione e streaming: neurale a entrambe le estremità, standard nel mezzo

Ciò che ricevono gli spettatori è ancora un bitstream standard

I maggiori servizi video hanno completato una transizione di codec e stanno avviando la successiva, e nessuna delle due coinvolge un decoder neurale. A dicembre 2025 Netflix ha comunicato che AV1 trasportava circa il 30% delle sue visioni, ed era il suo secondo codec più usato dopo AVC. AV1 è arrivato sui televisori solo quando ci sono arrivati i decoder hardware: Netflix ha aggiunto AV1 alla certificazione dei dispositivi nel 2019, lo ha trasmesso in streaming ai televisori dalla fine del 2021 e afferma che l'88% dei dispositivi a grande schermo presentati per la certificazione tra il 2021 e il 2025 lo supportava, quasi tutti dal 2023. YouTube ha dichiarato che al 19 marzo 2026 più del 90% del suo catalogo, ponderato per tempo di visione, era disponibile in AV1.

AV2, il successore di AV1 sviluppato dall'Alliance for Open Media, ha raggiunto la versione 1.0.0 il 28 maggio 2026, con una compressione migliore di circa il 30% rispetto ad AV1 secondo le misure oggettive di AOMedia. YouTube prevede stream AV2 sul web «già all'inizio del 2027». Il silicio è seguito subito: Chips&Media ha comple-

tato un decoder hardware AV2 a maggio e a giugno lo ha concesso in licenza a una grande azienda tecnologica nordamericana di cui non ha fatto il nome; Allegro DVT aveva consegnato la sua IP di decoder AV2 a tre clienti entro il 16 giugno; VeriSilicon aveva rilasciato un'IP video compatibile con AV2 a diversi clienti entro il 10 giugno. AV2 è un codec classico: i suoi guadagni vengono da partizionamento, predizione, trasformate e filtri in-loop migliori (uno dei quali è un filtro guidato di dettaglio «basato sui dati»), non da una rete end-to-end.

VVC, l'attuale standard ITU/ISO, ha trovato la sua prima applicazione broadcast in Brasile, dove il sistema DTV+ (TV 3.0) è entrato in esercizio commerciale a giugno 2026, in tempo per i Mondiali di calcio. A luglio 2026 nessun televisore in vendita in Brasile aveva un ricevitore DTV+ integrato; i primi sono attesi tra la fine del 2026 e l'inizio del 2027.

All'encoder: il downscaler neurale di Netflix

L'esempio più pulito di codifica neurale nella riproduzione è in funzione dal 2022 e non cambia nulla di ciò che deve fare il dispositivo dello spettatore. Netflix codifica ogni titolo a diverse risoluzioni, e la fase di downscaling che produce 1080p, 720p e 540p da un master 4K usava tradizionalmente un filtro fisso come Lanczos. Netflix lo ha sostituito con una piccola rete neurale, un blocco di pre-elaborazione seguito da un blocco di ridimensionamento, addestrata a migliorare la qualità visibile dopo un upscaler standard sul dispositivo. Netflix la descrive come una soluzione a sostituzione diretta: nessun'altra modifica né sul lato della codifica né sul lato client, per cui «milioni di dispositivi che supportano lo streaming di Netflix ne beneficiano automaticamente». Nei test di preferenza di Netflix circa il 77% degli spettatori ha scelto l'output del downscaler neurale **dato del produttore**. Per mantenere il costo accettabile alla scala di Netflix, «sono bastati pochi strati di rete neurale».

È lo schema da ricordare. La rete viene eseguita una volta per titolo nel data center, dove il calcolo costa poco e si può pianificare, e il suo output è un normale flusso AV1 o HEVC.

Sullo schermo: super-risoluzione nell'app e nel televisore

All'altra estremità, le reti riparano ciò che il codec e la sorgente hanno perso. Il 29 ottobre 2025 YouTube ha annunciato Super Resolution per la sua app per TV: i video caricati sotto il 1080p, da 240p a 720p, vengono portati in HD con l'apprendimento automatico, con il 4K come obiettivo dichiarato. Le versioni sottoposte a upscaling sono etichettate, gli spettatori possono tornare all'originale e i creator possono scegliere di esserne esclusi. A ottobre 2025 Meta ha descritto come esegue la super-risoluzione video sia sui suoi server al momento del caricamento, su normali CPU x86 in collaborazione con Intel, sia sul client in riproduzione, e ha riscontrato che le due cose si completano a vicenda. I produttori di televisori fanno lo stesso nel silicio: il processore 8K di punta di Samsung, l'NQ8 AI Gen3, dichiara 768 reti neurali per l'upscaling.

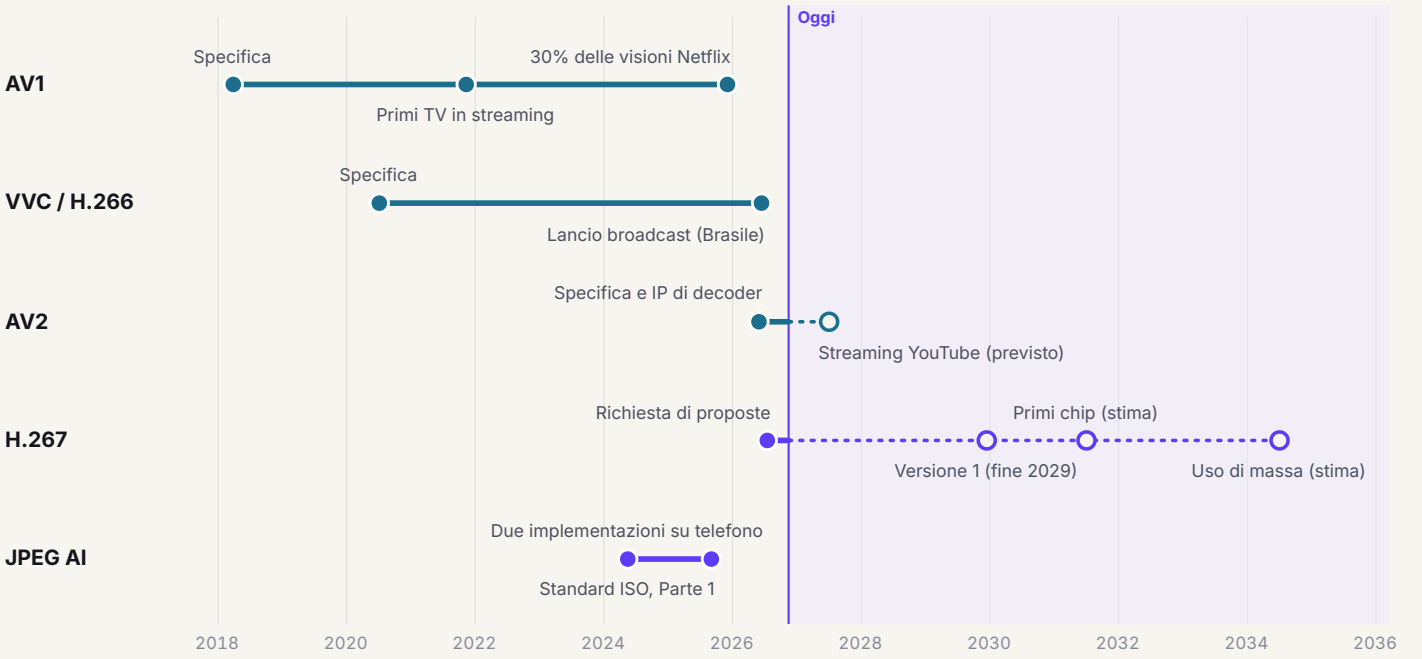
L'economia cambia a seconda di dove gira la rete. L'elaborazione lato server costa alla piattaforma potenza di calcolo per ogni video che tocca. L'elaborazione sul dispositivo non costa nulla alla piattaforma: usa una NPU che l'acquirente ha già pagato. Per questo i guadagni maggiori nel 2026 vengono da entrambe.

Perché nessun televisore decodifica un bitstream neurale

Un formato di riproduzione deve essere decodificabile su ogni dispositivo che lo riprodurrà mai, e quei dispositivi durano anni. Questo rende lento il percorso da uno standard finalizzato agli schermi di largo consumo, anche per i codec classici. La specifica di AV1 è stata finalizzata a marzo 2018; il decoder software dav1d è seguito nel giro di pochi mesi; i televisori lo hanno ricevuto in streaming dalla fine del 2021; ha raggiunto circa il 30% delle visioni su Netflix a dicembre 2025, quasi otto anni dopo la specifica. VVC è stato finalizzato a luglio 2020 e ha raggiunto il suo primo servizio broadcast a giugno 2026.

FIG. 03 **Dallo standard finalizzato agli schermi di largo consumo: buona parte di un decennio**

La riproduzione richiede un decoder hardware nel dispositivo. AV1 ha impiegato quasi otto anni dalla specifica finalizzata al 30% delle visioni su Netflix; il primo standard che probabilmente porterà strumenti neurali è previsto per la fine del 2029.



Fonte: Netflix Tech Blog (dic. 2025); AOMedia; Allegro DVT; TV Technology; Fraunhofer IIS sul DTV+ brasiliano; MPEG-155; comitato JPEG; ISO. I segni vuoti indicano obiettivi o stime dell'autore.

Un bitstream neurale aggiunge due problemi che i codec classici hanno risolto da tempo: l'energia per pixel e la decodifica esatta al bit. Entrambi sono trattati nel capitolo 8. Insieme significano che nessun produttore di televisori o telefoni metterà un decoder video neurale in ogni dispositivo prima che uno standard gli dica esattamente che cosa costruire.

Che cosa lo cambierebbe: H.267

Lo standard che potrebbe portare strumenti neurali su ogni schermo è il successore di VVC, sviluppato congiuntamente da ITU-T e ISO/IEC nel Joint Video Experts Team (JVET). Il nome non è definitivo; lo chiamerò H.267, come fa buona parte del settore.

A ottobre 2025 JVET ha valutato le risposte a una richiesta di evidenze: cinque risposte da 16 aziende e laboratori di ricerca, con 20 configurazioni di codec testate. Il comitato ha trovato evidenze di una compressione migliore di VVC e ha giudicato «interessanti» sia la tecnologia convenzionale migliorata sia quella basata su reti neurali. Gli approcci neurali hanno mostrato chiari guadagni percettivi, in particolare sui contenuti 8K HDR, e guadagni minori sui contenuti di gioco. InterDigital, uno dei rispondenti, ha presentato un progetto basato sull'Enhanced Compression Model (ECM) convenzionale e uno che sovrapponeva strumenti di IA a VVC.

La richiesta formale di proposte è stata emessa alla riunione di luglio 2026 a Ginevra. Le proposte principali andavano consegnate entro la fine di ottobre 2026, un pacchetto supplementare è atteso per metà dicembre e le risposte saranno valutate alla riunione JVET di Brisbane a gennaio 2027. JVET prevede un primo modello di test nella seconda metà del 2027 e una prima versione dello standard entro la fine del 2029, e definisce l'impresa «probabilmente uno sforzo più grande di qualunque altro mai compiuto per avviare un nuovo standard».

Il filone classico pone un'asticella alta. ECM, costruito con l'elaborazione convenzionale del segnale, risparmia all'incirca il 28% rispetto a VVC in configurazione random access. L'esplorazione di JVET sulle reti neurali riporta filtri in-loop che risparmiano circa il 6% a 10 mila moltiplicazioni-accumulo per pixel, l'8% a 22 mila e il 14% in un punto operativo ad alta complessità. La mia lettura è che H.267 sarà molto probabilmente un nucleo classico con pochi strumenti neurali, dove i bit risparmiati giustificano il costo del decoder, e che la valutazione di gennaio 2027 mostrerà quanti analisi.

Esiste già un ponte. I flussi VVC e HEVC possono trasportare informazioni supplementari di miglioramento (SEI) che descrivono un post-filtro neurale da applicare lato decoder, e le nuove edizioni di VVC e del suo standard SEI (VSEI), finalizzate a ottobre 2025, hanno esteso quei messaggi. Un flusso può indicare quale rete eseguire dopo la decodifica senza obbligare ogni decoder a eseguirla.

Che cosa significa per le aziende della riproduzione

Conviene vendere ora ciò che funziona attorno al codec: pre-elaborazione lato codifica per i servizi, e blocchi di post-elaborazione e taratura per i chip di televisori e set-top box. Non conviene costruire un'attività nella riproduzione su un bitstream neurale che debba essere decodificato su dispositivi di terzi prima degli anni 2030. Per chi produce silicio, i ricavi a breve termine sono i decoder AV2 e VVC; la parte neurale è un percorso NPU per il miglioramento accanto a essi (capitolo 8).

CAPITOLO 03

FOTOGRAFIA E IMMAGINI

03 Fotografia e immagini: neurale dentro la fotocamera, classico sul web

La pipeline della fotocamera è già neurale

Ogni foto scattata da un recente smartphone di punta è passata attraverso reti neurali prima di essere compressa. Lo Snapdragon 8 Gen 2 di Qualcomm, negli smartphone del 2023, ha introdotto un «Cognitive ISP» che esegue la segmentazione semantica in tempo reale durante la ripresa di foto e video, regolando separatamente volti, capelli, cieli e sfondi; il Galaxy S23 Ultra di Samsung è stato il primo telefono a usarlo sulla fotocamera per i selfie. A gennaio 2026 Chips&Media e Visionary.ai hanno presentato al CES quello che definiscono il primo ISP completo basato sull'IA: una pipeline d'immagine end-to-end che gira interamente su una NPU, la WAVE-N di Chips&Media, in virgola mobile a 16 bit, senza un ISP a funzione fissa, e che può essere aggiornata da remoto. «Le pipeline esistenti agganciano blocchi neurali a un ISP a funzione fissa. Noi sostituiamo completamente l'ISP convenzionale», ha detto il CEO di Visionary.ai. Dopo la fiera è stato offerto in valutazione ai produttori.

Poi il risultato viene salvato in JPEG o HEIF, formati classici che qualsiasi dispositivo può aprire. Lo schema è lo stesso della riproduzione: neurale dentro la pipeline di una singola azienda, standard al confine.

JPEG AI: uno standard neurale finalizzato, ancora senza prodotti

JPEG AI è il primo standard di codifica appresa delle immagini di ISO/IEC e ITU-T (ISO/IEC 6048, ITU-T T.840). La Parte 1, il sistema di codifica di base, è stata pubblicata nel 2025. Il comitato JPEG riporta guadagni di compressione dal 12,5% al 27,9% rispetto alla codifica intra di VVC, su una gamma di complessità del decoder da 7 a 216 mila moltiplicazioni-accumulo per pixel. Il suo progetto prevede due encoder e tre decoder, addestrati congiunta-

mente, così che un ricevitore possa implementare un solo decoder e leggere comunque ogni flusso. Entro maggio 2024 il comitato riportava due implementazioni su smartphone. In parallelo, a dicembre 2024 l'IEEE ha pubblicato IEEE 1857.11-2024, uno standard per la codifica di immagini basata su reti neurali progettato per le NPU.

A novembre 2026 non ho trovato alcuna fotocamera di telefono, app di foto o browser che scriva o legga JPEG AI di default.

Che cosa ha adottato invece il web: JPEG XL

Il formato d'immagine che ha guadagnato terreno nel 2026 è classico. Il 6 ottobre 2026 Google ha attivato di default la decodifica JPEG XL in Chrome 155, con un nuovo decoder memory-safe scritto in Rust, quattro anni dopo averne rimosso il supporto sperimentale. Safari decodifica JPEG XL dal 2023, e Mozilla si sta muovendo nella stessa direzione. Gli argomenti di Google a favore sono una compressione migliore di JPEG, la ricompressione senza perdita dei file JPEG esistenti, l'HDR e la decodifica progressiva.

La lezione vale in generale. Un formato che deve aprirsi ovunque vince con un decoder veloce e deterministico che gira su qualsiasi CPU e può essere reso memory-safe. Un decoder d'immagine neurale ha bisogno di una NPU o di una GPU e di un'aritmetica esatta su tutti i dispositivi. La codifica neurale vince dentro la pipeline di un dispositivo e nei sistemi in cui la stessa azienda codifica e decodifica, come un servizio di foto nel cloud che non consegna mai il file compresso a nessun altro.

Che cosa significa per le aziende dell'imaging

Il denaro è nella pipeline di acquisizione: IP di ISP con IA, NPU dimensionate per l'imaging e il lavoro di taratura che i produttori di fotocamere acquistano. Per l'archiviazione e la distribuzione, un codec d'immagine appreso è oggi uno strumento per sistemi chiusi, non un formato su cui scommettere.

CAPITOLO 04



CHIAMATE E RIUNIONI

04 Chiamate e riunioni: dove i codec neurali end-to-end si usano davvero

L'audio è arrivato per primo

Microsoft ha presentato Satin, un codec vocale basato sull'IA, a febbraio 2021, e ha detto che era già usato per tutte le chiamate tra due partecipanti in Teams e Skype. Satin offre parlato a banda super-larga, fino a 16 kHz di banda audio, a 6 kbps, mentre a quel bitrate il precedente codec Silk di Skype era a banda stretta. Il bitrate risparmiato viene speso in ridondanza, così le chiamate resistono meglio alla perdita di pacchetti.

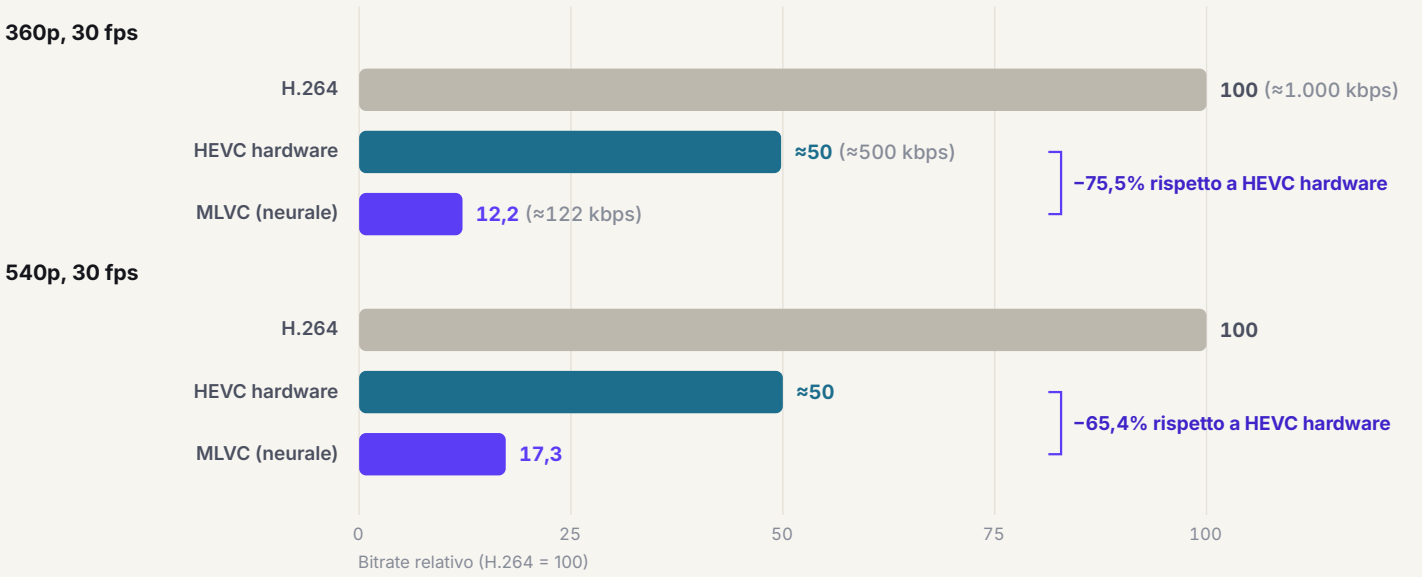
MLVC: un codec video neurale nelle chiamate reali

Il 24 luglio 2026 Microsoft ha rilasciato MLVC, il suo codec video basato sull'apprendimento automatico, come open source con licenza MIT, con il codice del modello, i pesi addestrati, gli script di addestramento e gli strumenti di conversione per le diverse NPU. MLVC è la versione di prodotto della famiglia DCVC che Microsoft Research pubblica dal 2021. Microsoft afferma di distribuire MLVC in Teams per le chiamate peer-to-peer, con telemetria in tempo reale, test A/B e i codec convenzionali come ripiego.

MLVC funziona a 540p e 360p a 30 fotogrammi al secondo ed è progettato per usare non più di metà della NPU sul Neural Engine di Apple e sulle NPU di Qualcomm e Intel, dividendo il lavoro tra CPU e NPU. Nei test soggettivi secondo ITU-T P.910 sul Video Conferencing Dataset di Microsoft, ha richiesto l'87,8% di bitrate in meno rispetto a H.264 e il 75,5% in meno rispetto a HEVC hardware a 360p, e l'82,7% e il 65,4% in meno a 540p dato del produttore. A 360p sono all'incirca 122 kbps, dove H.264 ha bisogno di circa 1 Mbps.

FIG. 04 **Nelle chiamate un codec neurale batte HEVC hardware del 65–76%**

Bitrate necessario a parità di qualità valutata sul set di test di videoconferenza di Microsoft, H.264 = 100. MLVC gira sulle NPU di portatili e telefoni fino a 540p30, usando al massimo metà della NPU.



Fonte: Microsoft, rilascio di MLVC (24 luglio 2026): test soggettivi ITU-T P.910 sul suo Video Conferencing Dataset. Barre HEVC ricavate dai risparmi dichiarati. Dati del produttore, non misure indipendenti.

Due cose ridimensionano il dato che fa notizia. L'88% è rispetto a H.264; rispetto all'HEVC hardware che gli stessi portatili contengono già, è il 65–76%, comunque molto. E MLVC ha rinunciato a parte della compressione per essere distribuibile: un'analisi indipendente osserva che un DCVC-RT addestrato su criteri percettivi ottiene l'81,9% sulla stessa misura, dove MLVC ottiene il 75,5%. Il prezzo della decodifica identica su hardware che Microsoft non possiede è stato di circa sei punti.

Perché le chiamate sono il terreno naturale

Una chiamata è un circuito chiuso. Entrambi gli endpoint eseguono la stessa applicazione, negoziano all'inizio della chiamata che cosa sono in grado di fare e possono ripiegare su un codec standard a metà chiamata. Le risoluzioni sono modeste, le reti sono spesso scadenti, e i portatili a entrambe le estremità ora sono dotati di NPU, come i telefoni da anni. Microsoft fissa a 40 TOPS di NPU la soglia per i suoi PC con IA, e il Neural Engine dell'M4 di Apple è accreditato di 38 TOPS. Ogni condizione che blocca un codec neurale nella riproduzione qui è assente.

I controesempi: Opus e MLow

Non tutte le piattaforme di chiamata hanno fatto la stessa scelta, e le ragioni sono istruttive.

Opus, il codec audio aperto di WebRTC e della maggior parte delle app di chiamata, ha introdotto il deep learning nella versione 1.5 (marzo 2024), il suo primo uso di reti per elaborare o generare il segnale stesso. Ma le ha introdotte come opzioni: un miglioramento lato decoder (LACE e NoLACE) che non richiede modifiche al bitstream, un

mascheramento neurale della perdita di pacchetti e uno schema di ridondanza profonda (DRED); sono tutti disattivati di default, così da non appesantire i dispositivi vecchi. Opus 1.6 ha aggiunto un'estensione di banda neurale e un modello DRED tre volte più piccolo, e resta pienamente compatibile con lo standard del 2012. L'approccio è «dentro uno standard, opzionale e retrocompatibile».

Meta ha fatto un passo ancora più indietro. A giugno 2024 ha lanciato MLow, un codec vocale a basso bitrate costruito sull'elaborazione classica del segnale anziché su una rete, ora in tutte le chiamate di Instagram e Messenger e in distribuzione su WhatsApp. Le ragioni di Meta erano i suoi dispositivi: più del 20% delle sue chiamate avviene su telefoni ARMv7, e decine di milioni di chiamate WhatsApp al giorno passano da telefoni con più di dieci anni. I codec neurali, compreso l'EnCodec della stessa Meta, funzionano in modo affidabile solo su telefoni costosi, ha scritto Meta. MLow offre circa il doppio della qualità di Opus a 6 kbps usando il 10% di calcolo in meno.

Per un servizio i cui utenti vanno dai portatili di punta ai telefoni vecchi di dieci anni, decide il budget di complessità. I codec neurali arrivano prima dove la base installata è nuova e il fornitore controlla entrambe le estremità.

Che cosa significa per le aziende delle comunicazioni

Videoconferenze aziendali, telepresenza, assistenza remota, collegamenti di contribuzione e reti private sono i contesti in cui oggi un codec neurale end-to-end è un prodotto, a patto di controllare entrambe le estremità. Occorre prevedere tre cose fin dall'inizio: un codec di ripiego, test di decodifica su ogni NPU supportata e una valutazione della qualità sui propri contenuti anziché sul dataset di un fornitore.

CAPITOLO 05

IA GENERATIVA

05 IA generativa: il più grande impiego che nessuno chiama codec

Ogni modello vocale e video si basa su un codec appreso

Il più grande impiego della compressione neurale end-to-end non è in alcuno standard di codifica. È dentro l'IA generativa. Un modello vocale non ascolta né produce direttamente forme d'onda: legge e scrive i token discreti di un codec audio appreso. Mimi di Kyutai, rilasciato nel 2024, trasforma il parlato in token a 12,5 frame al secondo e 1,1 kbps, meno di qualsiasi codec telefonico, combinando informazioni semantiche e acustiche in un unico flusso. EnCodec di Meta, pubblicato nel 2022, è un altro esempio; gli stessi ingegneri di Meta hanno poi descritto codec di questo tipo come eccellenti a bassi bitrate ma troppo pesanti per i telefoni più vecchi (capitolo 4).

La generazione video funziona allo stesso modo. Un modello di diffusione o un transformer genera nello spazio latente di un autoencoder video appreso (VAE), e il decoder del VAE trasforma il risultato in pixel. Il VAE di LTX-Video comprime il video 1:192: ogni token latente rappresenta un blocco di 32×32 pixel su 8 fotogrammi, con 128 canali. Il suo decoder esegue anche l'ultimo passo di denoising, quindi in parte genera i dettagli invece di limitarsi a ricostruirli.

Perché non sono formati di distribuzione

Questi codec girano a ogni richiesta rivolta a un modello vocale o video, ma nessuno di essi è un formato che qualcun altro decodifichi. Ogni modello ha il proprio spazio latente; non c'è controllo del bitrate, né test di

conformità, né garanzia che un decoder diverso produca la stessa immagine. Sono progettati per la qualità di generazione, non per la fedeltà a un dato bitrate, e decodificano su GPU di data center. Il loro successo dimostra che la compressione appresa funziona su larga scala quando una sola azienda possiede l'encoder, il decoder e l'hardware.

Che cosa significa per le aziende

Due punti pratici. Primo, l'efficienza del tokenizer è una voce di costo diretta per chiunque eroghi video o voce generativi: per un transformer che applica l'attenzione su tutti i token, il calcolo cresce più rapidamente del numero di token, quindi una rappresentazione latente più compressa riduce il costo di generazione più che proporzionalmente analisi. Gli ingegneri dei codec che conoscono la progettazione dello spazio latente, i modelli entropici e l'efficienza dei decoder sono preziosi per i team di IA generativa. Secondo, il confine si sta spostando: gli enti di standardizzazione hanno avviato lavori sulla codifica per le macchine anziché per le persone. Video Coding for Machines di MPEG (ISO/IEC 23888-2) ha raggiunto lo stadio di bozza di standard internazionale all'inizio del 2026 e la versione finale è prevista per la fine del 2026, e JPEG AI è stato progettato fin dall'inizio per servire compiti di visione artificiale oltre alla visione umana.

CAPITOLO 06

VIDEOGIOCHI E GRAFICA

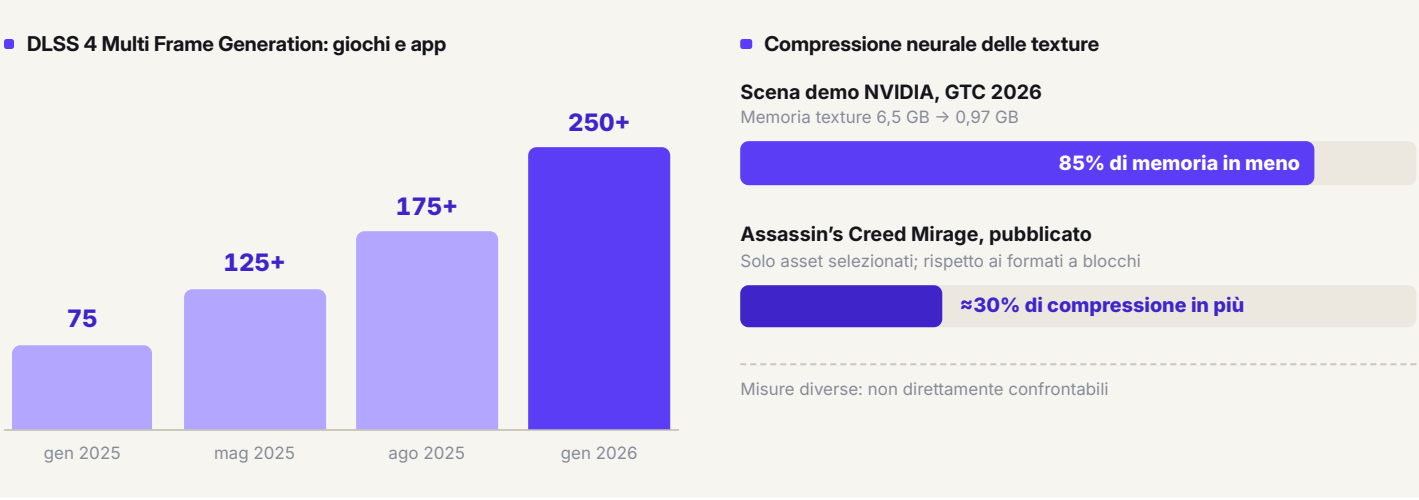
06 Videogiochi e grafica: la ricostruzione neurale è diffusa, gli asset neurali sono agli inizi

Renderizzare meno pixel e lasciare che una rete completi il resto

Nella grafica in tempo reale una rete ricostruisce già la maggior parte di ciò che i giocatori vedono. Il DLSS di NVIDIA renderizza meno pixel e meno fotogrammi e usa reti per produrre il resto. DLSS 4 con Multi Frame Generation è stato lanciato a gennaio 2025 con il supporto di 75 giochi e app; NVIDIA ne contava più di 125 a maggio 2025, più di 175 ad agosto e più di 250 a gennaio 2026, quando ha annunciato DLSS 4.5 con un modello transformer di seconda generazione per la super-risoluzione. È un decoder a tutti gli effetti, tranne che nel nome: il «bitstream» è un fotogramma a bassa risoluzione più i dati di movimento, e la rete ricostruisce l'immagine completa.

FIG. 05 **Quando un'unica azienda controlla l'intero stack, la ricostruzione neurale si diffonde in fretta**

A sinistra: giochi e app che supportano DLSS 4 Multi Frame Generation di NVIDIA. A destra: compressione neurale delle texture in una scena dimostrativa e in un gioco pubblicato, che la usa solo su asset selezionati per il suo costo in esecuzione.



Fonte: NVIDIA (da gennaio 2025 a gennaio 2026; demo al GTC 2026); Ubisoft La Forge (febbraio 2026). Dati del produttore; i due valori sulle texture misurano cose diverse.

Le texture neurali sono arrivate, in modo selettivo

Le texture sono compresse in formati a blocchi fissi (BCn) che le GPU decodificano in hardware. La compressione neurale delle texture sostituisce un insieme di texture dei materiali con una rappresentazione latente compatta e una piccola rete che ricostruisce i texel quando servono. A febbraio 2026 il laboratorio La Forge di Ubisoft ha descritto come l'ha portata in Assassin's Creed Mirage, un gioco del 2023, con una compressione effettiva superiore di circa il 30% rispetto ai formati convenzionali a parità di qualità visiva dato del produttore. Non è stata usata ovunque: «l'uso era limitato dalle prestazioni in esecuzione», per cui Ubisoft l'ha applicata a un sottoinsieme di asset con molte istanze e una forte pressione sulla memoria.

L'RTX Neural Texture Compression SDK di NVIDIA, in beta su GitHub, mostra il limite superiore. In una dimostrazione al GTC 2026, una scena di villa che usava 6,5 GB di memoria per le texture con BCn ne richiedeva 970 MB con le texture neurali, l'85% in meno dato del produttore. L'SDK offre due percorsi: decodificare in BCn al caricamento, il che fa risparmiare spazio su disco e dimensione del download ma non memoria video, oppure decodificare al momento del campionamento, il che fa risparmiare la memoria ma costa tempo di shader a ogni fotogramma.

L'API sta arrivando

Le API grafiche stanno aggiungendo le operazioni matriciali di cui la decodifica neurale ha bisogno dentro gli shader. Microsoft ha rilasciato Shader Model 6.9 a febbraio 2026; la funzione cooperative vector presentata in anteprima nel 2025 è confluita in un progetto unificato di algebra lineare per un futuro shader model, e DirectX Linear Algebra è entrato in anteprima pubblica ad aprile 2026. Quando sarà presente in ogni driver, un decoder neurale di texture potrà girare sulla GPU di qualsiasi produttore da un unico percorso di codice.

Che cosa significa per le aziende

I giochi adottano la decodifica neurale perché valgono le condizioni di un circuito chiuso: un solo motore, una sola pipeline degli asset, hardware tensoriale nella GPU e un formato di ripiego. Il lavoro aperto riguarda gli strumenti, le pipeline degli asset che codificano materiali neurali, i runtime indipendenti dal produttore e, soprattutto, il mobile, dove a decidere è il costo energetico per texel decodificato.

CAPITOLO 07

07

COLLEGAMENTI SPECIALIZZATI

Collegamenti specializzati: satelliti, droni e robot

Sono i casi in cui la codifica neurale end-to-end si adatta meglio, sulla carta. Un unico proprietario controlla entrambe le estremità del collegamento, la banda è scarsa e soggetta a perdite, il contenuto è abbastanza circoscritto da poterci addestrare una rete, e spesso basta un'immagine fedele sul piano percettivo o persino soltanto su quello semantico. I codec neurali resistenti alle perdite per questi collegamenti sono un'area di ricerca attiva, e il restauro lato server di flussi a basso bitrate è già in uso.

Sono anche un mercato piccolo e sempre più regolamentato rispetto al video di consumo analisi. In Cina, l'Annuncio n. 49 del 2026 del Ministero del Commercio e di altri due enti, emesso il 19 ottobre e in vigore dal 18 novembre 2026, rivede i controlli all'esportazione su 13 categorie di articoli legati ai droni. Tra le voci tecnologiche, una riguarda le piattaforme terrestri senza equipaggio definite in parte da soglie come una compressione semantica delle immagini del 75% e una ricostruzione dell'immagine entro 150 millisecondi, secondo quanto riportato dalla stampa di settore. La compressione semantica stessa è diventata un parametro dei controlli all'esportazione. Negli Stati Uniti, il 22 dicembre 2025 la Federal Communications Commission ha aggiunto alla sua Covered List i droni prodotti all'estero e i loro componenti critici, per cui i nuovi modelli stranieri non possono più ottenere l'autorizzazione delle apparecchiature.

Il verdetto: una nicchia solida per un'azienda specializzata con clienti industriali o governativi e la capacità di gestire la conformità alle regole sull'export da entrambe le parti, ma non un mercato di massa, né il luogo in cui capire se la codifica neurale funziona su larga scala.

CAPITOLO 08

SILICIO

08 La prospettiva del silicio: energia, calcolo, memoria ed esattezza

Questo è il capitolo in cui la mia esperienza conta di più. Ho alle spalle molti anni di R&S e di gestione nel settore dei chip fabless, tra cui la progettazione full-custom di datapath ASIC e progetti di chip multimediali per fotocamere come acceleratori di codec video, e le domande che un team di chip pone su un nuovo codec non riguardano il BD-rate. Sono queste: quanti joule per pixel, quanti byte di traffico DRAM per pixel, quante porte logiche, e ogni decoder al mondo può calcolare esattamente gli stessi bit? Su tutti e quattro i punti, un decoder neurale nel 2026 è dove era HEVC hardware prima dei suoi primi chip: promettente nel software, costoso nel silicio.

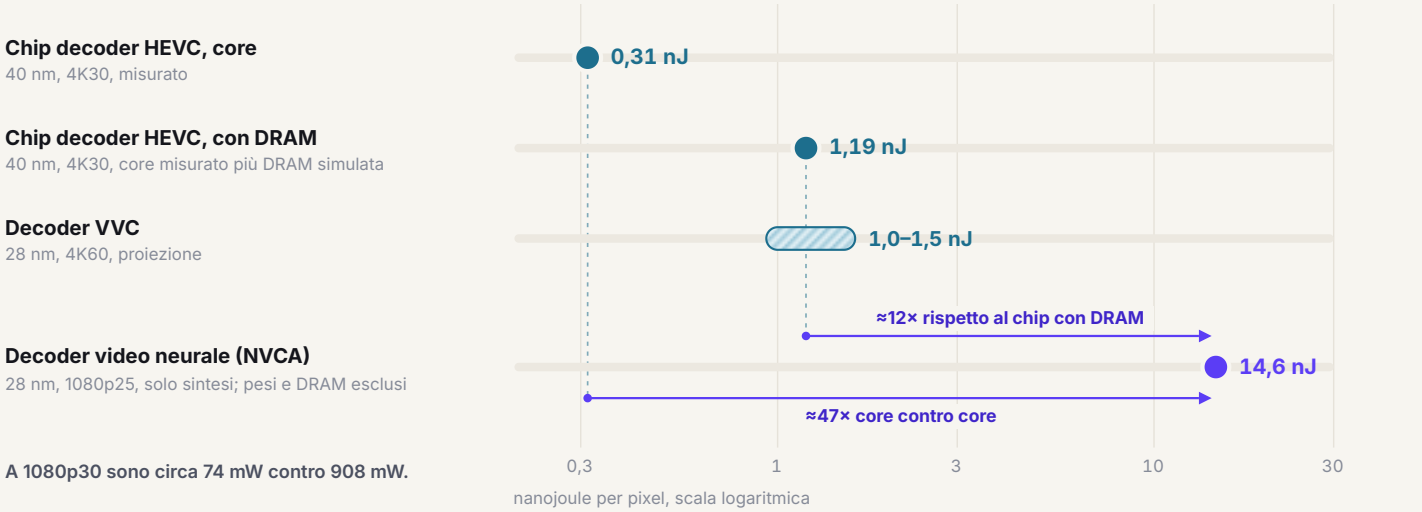
Energia per pixel: un ordine di grandezza di distanza

Pochi decoder neurali sono stati progettati come chip, quindi il confronto si basa su una manciata di progetti pubblicati. Il riferimento sul lato classico è il chip decoder HEVC del MIT (Tikekar et al., ISSCC 2013 e IEEE JSSC 2014): 40 nm, 1,77 mm², decodifica del 4K a 30 fotogrammi al secondo a 200 MHz. Il suo core consuma 78 mW, ovvero 0,31 nJ per pixel; conteggiando la potenza simulata della DRAM, il sistema richiede 1,19 nJ per pixel. Per VVC i dati di produzione sono proprietari; un'analisi pubblicata su Electronic Design nel 2026, che scala il chip del MIT a 28 nm e applica la maggiore complessità di VVC, prevede 1–1,5 nJ per pixel per un decoder 4K60 di circa un milione di porte logiche.

Sul lato neurale, l'unico progetto ASIC pubblicato di decoder video neurale end-to-end, secondo il conteggio di Electronic Design, è NVCA (Zhang et al., DATE 2024), sintetizzato in TSMC 28 nm con 5,01 milioni di porte logiche, senza contare i pesi, che risiedono fuori dal chip. Decodifica il 1080p a 25 fotogrammi al secondo a 760 mW, ovvero 14,6 nJ per pixel.

FIG. 06 **Energia per pixel decodificato: i decoder a funzione fissa sono avanti di un ordine di grandezza**

L'unico progetto pubblicato di chip decoder video neurale richiede circa 12 volte l'energia per pixel di un chip decoder HEVC misurato, conteggiato con la sua DRAM, e circa 47 volte quella del solo core. Il traffico dei pesi e della DRAM del progetto neurale non è conteggiato.



Fonte: Tikekar et al., IEEE JSSC 2014 (MIT); NVCA, Zhang et al., DATE 2024 (arXiv 2312.10716); l'intervallo VVC è una proiezione di Electronic Design (2026). Il chip HEVC usa un nodo più vecchio (40 nm contro 28 nm), il che favorisce il progetto neurale.

Sono circa 12 volte il chip HEVC conteggiato con la sua DRAM, e circa 47 volte il suo solo core. Il confronto favorisce NVCA in tre modi: il chip HEVC è su un nodo più vecchio, il dato di NVCA esclude il traffico dei pesi e della DRAM, e NVCA è un risultato di sintesi anziché silicio misurato. A 1080p30 la differenza è di circa 74 mW contro all'incirca 900 mW: un decoder che nessuno nota, contro uno che pesa sulla batteria e sul budget termico di un telefono.

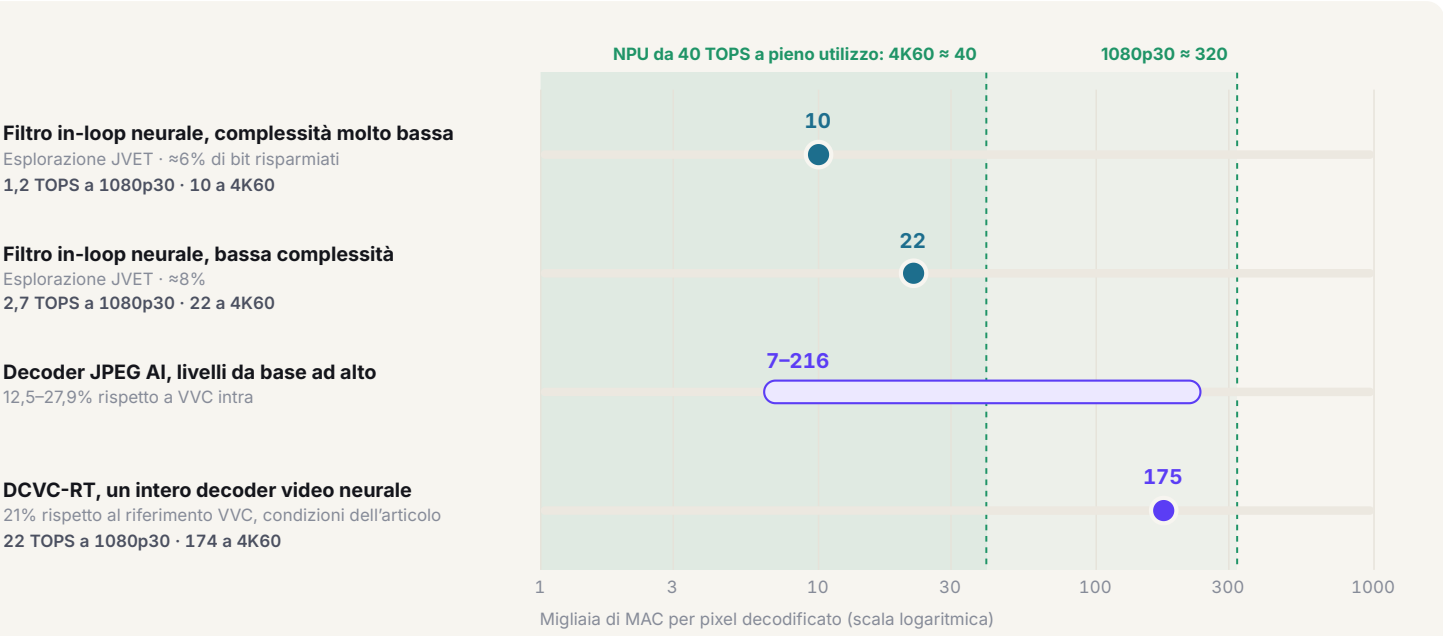
Calcolo per pixel: che cosa chiede un decoder neurale a una NPU

Il codec video neurale pubblicato più veloce che ho trovato è DCVC-RT di Microsoft Research (Jia et al., CVPR 2025). Codifica e decodifica il 1080p a 125 e 113 fotogrammi al secondo su una GPU NVIDIA A100, risparmiando il 21% di bitrate rispetto all'encoder di riferimento di VVC nelle condizioni di test dell'articolo, e usa l'aritmetica intera così che dispositivi diversi decodifichino in modo coerente. Il suo materiale supplementare elenca la complessità di ciascun modulo. Sommando i quattro moduli che un decoder deve eseguire (decoder, estrattore di feature, modello entropico e ricostruzione), ottengo circa 175 mila moltiplicazioni-accumulo per pixel decodificato stima.

A 1080p30 sono circa 11 mila miliardi di moltiplicazioni-accumulo al secondo, ovvero circa 22 TOPS a due operazioni per MAC, prima di qualsiasi perdita di utilizzo; il 4K60 ne richiede otto volte tanto stima. Per dare un'idea della scala, Microsoft definisce i suoi PC con IA con una NPU di almeno 40 TOPS, e Apple accredita il Neural Engine dell'M4 di 38 TOPS. Sono valori di picco, condivisi con tutto ciò che il dispositivo fa.

FIG. 07 **Uno strumento neurale costa decine di migliaia di MAC per pixel; un intero decoder neurale centinaia di migliaia**

Moltiplicazioni-accumulo del decoder per pixel, in scala logaritmica, rispetto a quanto potrebbe permettersi una NPU da 40 TOPS a pieno utilizzo a 4K60 e a 1080p30. I guadagni sono misurati rispetto a riferimenti diversi e non sono direttamente confrontabili.



Fonte: JVET tramite il presidente di SC 29 (aprile 2026); comitato JPEG (2024); DCVC-RT, materiale supplementare CVPR 2025, somma dell'autore dei moduli del decoder; definizione di PC Copilot+ di Microsoft. Calcoli dell'autore: due operazioni per MAC, utilizzo pieno.

La progressione della figura 7 è il modo utile di leggere il settore. Gli strumenti neurali che costano decine di migliaia di MAC per pixel, come i filtri in-loop di JVET (circa il 6% di risparmio a 10 mila, l'8% a 22 mila), sono plausibili dentro un futuro decoder standard. JPEG AI copre l'intera gamma, da 7 a 216 mila MAC per pixel per un 12,5–27,9% su VVC intra. Un intero decoder video neurale a un paio di centinaia di migliaia di MAC per pixel sta su una NPU, a risoluzione modesta, in un circuito chiuso. È esattamente lì che si colloca MLVC: 540p30, usando al massimo metà della NPU.

Il traffico di memoria, non i MAC, è spesso il vero limite

Gli autori di DCVC-RT hanno scoperto che a limitare la velocità non era l'aritmetica ma il «costo operativo»: l'input e l'output di memoria e il numero di chiamate di funzione. Hanno eliminato i moduli espliciti di movimento e l'elaborazione multiscala in buona parte per questo motivo. I progettisti di NVCA hanno costruito un flusso di dati apposito per ridurre il traffico verso la memoria esterna al chip dovuto alle feature di movimento e di residuo. Un progettista di chip riconoscerà lo schema. Nel chip HEVC del MIT la DRAM rappresenta circa tre quarti dell'energia del sistema (0,88 su 1,19 nJ per pixel), anche dopo una cache che ha ridotto la banda DRAM del 67% analisi. Un decoder neurale che mantiene grandi mappe di feature multicanale da un fotogramma all'altro peggiora il problema della memoria, non lo attenua.

Esattezza: ogni decoder deve calcolare gli stessi bit

La codifica entropica funziona solo se il modello di probabilità del decoder coincide esattamente con quello dell'encoder. In un codec classico quel modello è una tabella o un piccolo insieme di regole su interi. In un codec neurale è una rete, e una differenza in virgola mobile tra due dispositivi, dovuta a un kernel diverso, a un ordine di

accumulo diverso o a una NPU diversa, può far leggere al decoder i simboli sbagliati e fargli perdere il flusso da quel punto in poi.

Chiunque abbia portato in produzione o standardizzato un codec neurale ha dovuto eliminare il problema con l'ingegneria. DCVC-RT converte il suo modello in interi per una codifica coerente tra dispositivi. MLVC usa un hyperprior a scala condivisa per resistere alle discrepanze tra piattaforme, il che è probabilmente parte del motivo per cui rinuncia a compressione rispetto a DCVC-RT analisi. L'esplorazione di JVET sulle reti neurali fa passare ogni rete attraverso SADL, una piccola libreria con inferenza su interi a 16 bit. JPEG AI definisce punti di conformità nei domini del latente, del residuo e dei pixel. Per un team di chip è la stessa disciplina delle trasformate intere di HEVC: lo standard deve definire l'aritmetica in modo esatto, il che vincola il progetto della rete ed esclude scorciatoie specifiche dei singoli produttori.

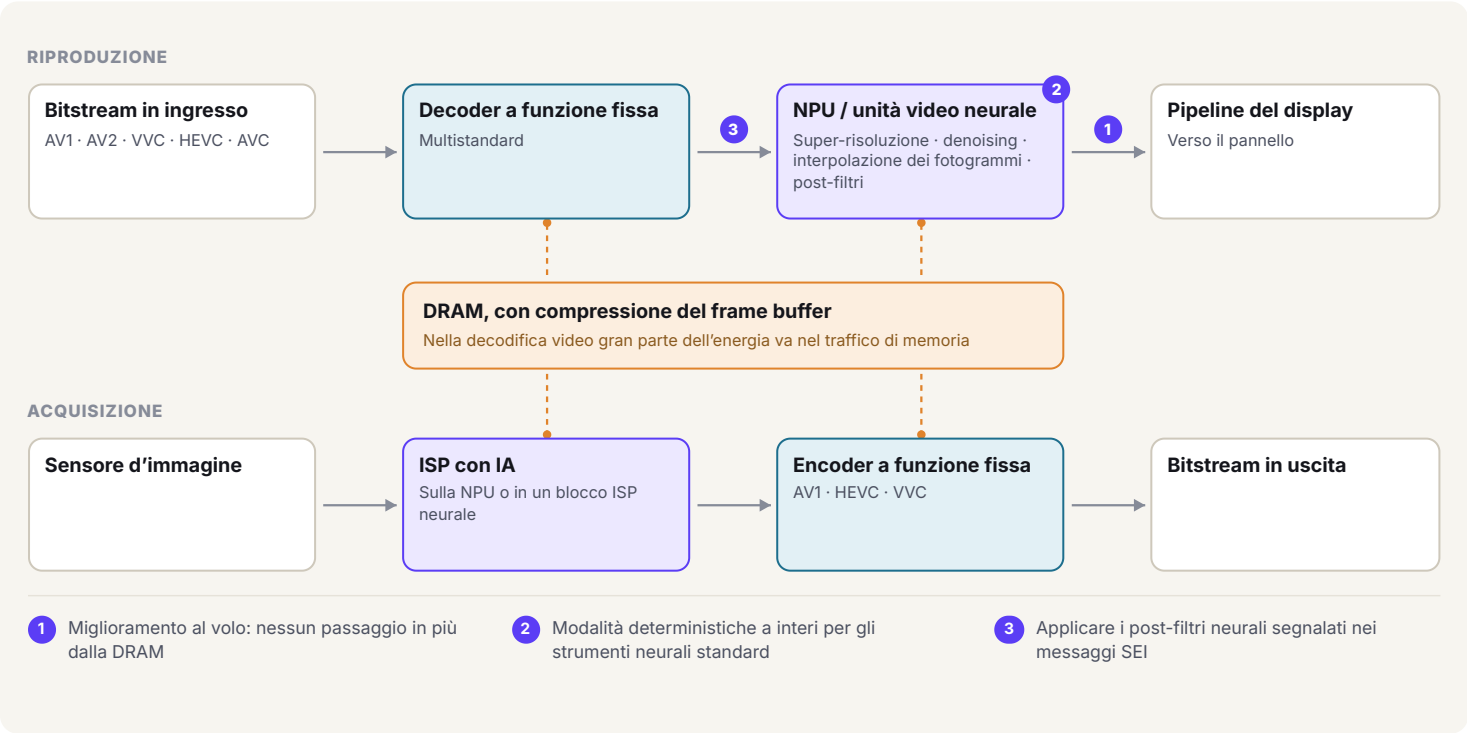
Il problema non è l'area, ma la verifica e il ciclo di vita

I 5 milioni di porte logiche di NVCA contro circa un milione per il decoder VVC della proiezione sembrano tanti, ma su un moderno SoC per telefoni o televisori che porta già una NPU da diversi TOPS, qualche milione di porte non è ciò che decide il progetto. Ciò che costa davvero è la verifica rispetto a una suite di conformità, mantenere corretto un decoder programmabile attraverso gli aggiornamenti software per tutta la vita del dispositivo, e la potenza a 4K60 HDR. Questi costi hanno senso solo una volta che uno standard fissa che cosa va costruito.

L'architettura che vince oggi: un codec a funzione fissa più una NPU

FIG. 08 **La ripartizione che arriva nei prodotti: hardware a funzione fissa per il bitstream, un motore neurale per i pixel**

Un SoC multimediale del 2026. Il blocco codec decodifica ogni standard con una frazione di nanojoule per pixel; la NPU fa ciò in cui le reti eccellono e che gli standard non vincolano. Il costo che conta è il traffico DRAM.



Fonte: Visione architetturale dell'autore, basata sulle informazioni di prodotto diffuse da Tencent (Canghai V2), Chips&Media, Allegro DVT e Samsung (2026).

Il silicio che si vende nel 2026 abbina un codec multistandard a funzione fissa a un motore neurale al suo fianco. Il blocco codec gestisce AV1, AV2, VVC, HEVC e AVC con una frazione di nanojoule per pixel. Il motore neurale

gestisce ciò in cui le reti eccellono e che gli standard non vincolano: il denoising e l'ISP in acquisizione, la super-risoluzione, l'interpolazione dei fotogrammi e il miglioramento in visualizzazione. I due condividono la compressione del frame buffer e la pianificazione della DRAM, perché il traffico di memoria è il costo che conta.

Il mercato mostra questa ripartizione. L'ASIC codec video Canghai (沧海) di Tencent, sviluppato internamente dal 2019, conta più di 100.000 chip di prima generazione installati in Tencent Cloud e nei servizi di streaming in diretta, on demand e di cloud gaming dell'azienda; la seconda generazione, annunciata a maggio 2026 come già accesa e in ingresso nella produzione in volumi per entrare in servizio nella seconda metà dell'anno, supporta H.265 e H.266 in un'unica architettura, con una compressione migliore rispettivamente di oltre il 10% e il 30% rispetto alla generazione precedente, circa il doppio del throughput e il miglioramento dell'immagine sul chip, come la super-risoluzione. Chips&Media vende fianco a fianco IP di codec e una NPU orientata al video, WAVE-N. Allegro DVT vende decoder AV2 e ha aggiunto la sua prima IP di elaborazione video neurale. I processori per TV di Samsung abbinano decoder standard a centinaia di reti di upscaling.

Cosa costruirei, per ruolo

Team di SoC e chip per TV. Mantenere decoder a funzione fissa per ogni standard da riprodurre, compreso AV2 ora e VVC dove lo richiede il broadcast. Dimensionare la NPU per la post-elaborazione alla risoluzione e alla frequenza di fotogrammi del display, e progettare il percorso dal decoder alla NPU al display in modo che il miglioramento non aggiunga un passaggio di andata e ritorno attraverso la DRAM. Supportare le SEI per post-filtri neurali. Prevedere modalità deterministiche a interi nella NPU. Considerare H.267 come un avvio di progetto nel 2029 o più tardi, e mantenere un po' di programmabilità per i suoi strumenti neurali.

Fornitori di IP. Vendere ora decoder AV2 e VVC, più i componenti che servono intorno a essi per l'elaborazione neurale: acceleratori di post-filtri, compressione del frame buffer e strumenti di conformità.

Piattaforme video. La pre-elaborazione neurale all'encoder ripaga per prima, perché gira una volta per titolo e non chiede nulla ai dispositivi. Usare la super-risoluzione lato client dove i dispositivi lo permettono. Riservare i codec neurali end-to-end ai circuiti chiusi sotto il proprio controllo.

CAPITOLO 09

STANDARD E BREVETTI

09 Standard e brevetti: dove gli strumenti neurali diventano oggetto di licenza

Gli standard in cantiere

Standard	Ente	Contenuto neurale	Stato, novembre 2026
H.267 (nome provvisorio; oltre VVC)	JVET (ITU-T e ISO/IEC)	Aperto. Strumenti classici (ECM) e strumenti neurali (filtri in-loop, predizione intra, super-risoluzione) sono entrambi candidati	Richiesta di proposte luglio 2026; risposte valutate gennaio 2027; primo modello di test seconda metà del 2027; versione 1 fine 2029
VVC e HEVC, con VSEI	JVET	Un post-filtro neurale può essere descritto nei metadati SEI; applicarlo è facoltativo per il decoder	In vigore; messaggi estesi nelle edizioni di ottobre 2025
AV2	AOMedia	Strumenti classici, esenti da royalty	Versione 1.0.0 il 28 maggio 2026; IP di decoder concessa in licenza da giugno 2026; stream web di YouTube previsti per inizio 2027
JPEG AI	JPEG (ISO/IEC e ITU-T)	Codec d'immagine appreso end-to-end, per la visione umana e artificiale	Parte 1 pubblicata nel 2025; una seconda versione in discussione; nessun supporto di default nei prodotti trovato
IEEE 1857.11	IEEE	Codifica di immagini basata su reti neurali progettata per NPU	Pubblicato a dicembre 2024
Video Coding for Machines	MPEG	Codifica per l'analisi automatica anziché per la visione	Bozza di standard internazionale a inizio 2026; versione finale prevista per fine 2026
Estensioni ML di Opus	IETF	Ridondanza profonda, miglioramento lato decoder ed estensione di banda, opzionali	In libopus 1.5 e 1.6; ridondanza profonda in fase di standardizzazione

Due cose spiccano. Primo, i contenuti neurali entrano negli standard ai margini prima che nel nucleo: come post-filtri opzionali, miglioramenti opzionali e codifica di immagini, tutti elementi che lasciano invariato il percorso principale di decodifica. Secondo, l'unico standard che potrebbe mettere strumenti neurali nel percorso di decodifica di ogni schermo, H.267, è a tre anni da una prima versione.

Chi si è posizionato sui brevetti

Gli strumenti di uno standard diventano brevetti soggetti a licenza, quindi il processo di H.267 è anche una corsa ai brevetti, e gli strumenti neurali ne fanno parte. Il panorama brevettuale di GreyB del 2025, su 3.916 brevetti post-VVC, mette in testa ByteDance, con 645 famiglie di brevetti e 251 contributi a JVET, e segnala partnership tra università e industria come quella tra Xiaomi e la Xidian University, con più di 20 famiglie. MLex ha riportato a marzo 2026 che Tencent, Alibaba e ByteDance stanno emergendo come partecipanti attivi nei pool di brevetti dei codec video. A maggio 2026 Sauvegarder Investment Management ha acquistato da Alibaba 126 asset brevettuali

sui codec video che coprono AV1, AV2 e Video Coding for Machines. L'acquisizione di Deep Render da parte di InterDigital, a ottobre 2025, ha aggiunto al suo portafoglio i brevetti di Deep Render sulla codifica video con IA.

Il denaro per dispositivo si vede facilmente nei tariffari dei pool. Le tariffe di Access Advance per VVC (al 1º luglio 2026, per licenziatari in regola che usano il suo marchio) sono di 0,50 dollari per telefono cellulare, tablet o portatile nei mercati della Region 1 come Stati Uniti, UE e Giappone, e di 0,25 dollari nella Region 2; per i televisori 4K, da 1,00 a 1,50 dollari a seconda del prezzo nella Region 1; con un tetto annuo per impresa di 75 milioni di dollari. AV1 e AV2 sono esenti da royalty da parte dei membri di AOMedia. Qualunque cosa contenga H.267, i suoi strumenti saranno tariffati nello stesso modo, ed è per questo che le aziende con brevetti sulla codifica neurale stanno contribuendo a JVET già ora analisi.

CAPITOLO 10

10 CHI GUADAGNA
Chi guadagna, e dove può inserirsi un'azienda

Il denaro nel 2026

FIG. 09 **Chi guadagna con la codifica neurale nel 2026**

Il denaro segue gli standard e le piattaforme. Non ho trovato nessuna azienda indipendente che viva della licenza di un codec video neurale end-to-end.

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)

Fonte: Informazioni diffuse da aziende e pool di brevetti citate nel rapporto; situazione al 12 novembre 2026. Tariffe del pool: Access Advance, VVC, Region 1, licenziatari in regola che usano il marchio, al 1º luglio 2026.

Le piattaforme si tengono i risparmi. Netflix, YouTube, Meta, Microsoft e Tencent sviluppano internamente l'elaborazione neurale e ne trattengono il beneficio sotto forma di meno banda, meno spazio di archiviazione e una qualità migliore per bit. Quando aiuta l'adozione regalano il codice: Microsoft ha rilasciato MLVC con licenza MIT.

I produttori di dispositivi e di chip vendono differenziazione. Gli ISP delle fotocamere e i processori d'immagine dei televisori sono il punto in cui le reti neurali toccano ogni utente, e sono venduti come funzionalità: il Cognitive ISP di Qualcomm, i processori di upscaling di Samsung.

I fornitori di IP vendono decoder per ogni nuovo standard, più NPU. Entro tre settimane dalla specifica di AV2, Chips&Media, Allegro DVT e VeriSilicon hanno tutti annunciato clienti per il decoder AV2. Gli stessi fornitori ora vendono blocchi di elaborazione neurale accanto ai codec.

I produttori di GPU costruiscono piattaforme. DLSS e la compressione neurale delle texture rendono l'hardware tensoriale di NVIDIA parte del modo in cui si fanno i giochi.

I titolari di brevetti incassano per dispositivo, attraverso i pool per VVC e HEVC e, dopo il 2029, per H.267.

Le startup di codec neurali sono state acquisite. WaveOne è andata ad Apple nel 2023, iSIZE a Sony Interactive Entertainment nel 2023 e Deep Render a InterDigital nel 2025, tutte a condizioni non rese note. In ogni caso una piattaforma o un licenziante ha acquistato il team e i brevetti. Nel 2026 non ho trovato nessuna azienda indipendente che viva della licenza di un codec video neurale end-to-end.

Dove può inserirsi un'azienda

Opportunità	Chi compra	Perché ora	Rischio principale
Pre-elaborazione neurale all'encoder (downscaling, denoising, taratura per titolo)	Piattaforme di streaming e di video degli utenti, broadcaster	Nessuna modifica necessaria ai dispositivi; collaudata da Netflix	Le grandi piattaforme se la costruiscono da sole
IP e taratura per il miglioramento lato display (super-risoluzione, interpolazione dei fotogrammi, qualità d'immagine)	Produttori di chip per TV, set-top box e telefoni	Le NPU sono già nei dispositivi; differenziazione visibile	Commoditizzazione; produttori di chip che sviluppano in proprio
Codec neurali a circuito chiuso (videoconferenze, operazioni a distanza, collegamenti privati)	Fornitori di comunicazioni aziendali, industriali e di telepresenza	MLVC dimostra la fattibilità, e il suo codice è aperto	Occorre controllare entrambe le estremità; frammentazione delle NPU
Silicio o IP codec più NPU (AV2, VVC e post-elaborazione)	Produttori di SoC	Il ciclo di AV2 sta iniziando	Operatori già affermati e forti
Strumenti di 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

Per la maggior parte delle aziende, le prime due righe sono quelle in cui i ricavi sono disponibili nel 2027. La terza è la più interessante per un'azienda di prodotto focalizzata. L'ultima richiede pazienza e una strategia di licenza.

CAPITOLO 11

11 AFFERMAZIONI E PROVE
Affermazioni ricorrenti, e che cosa dicono le prove

Affermazione	Che cosa dicono le prove
«I codec neurali battono già VVC»	In laboratorio, sì. DCVC-RT risparmia il 21% rispetto all'encoder di riferimento di VVC a 1080p nelle condizioni del suo articolo; JPEG AI risparmia il 12,5–27,9% rispetto a VVC intra. Il decoder costa circa 175 mila MAC per pixel, ovvero l'energia di una NPU, contro una frazione di nanojoule nell'hardware a funzione fissa.
«I codec IA sostituiranno H.26x e AV1 entro pochi anni»	Non nella riproduzione. Il primo standard in grado di portare strumenti neurali punta alla fine del 2029, i chip seguiranno verso il 2031 e l'uso di massa a metà degli anni 2030 stima. Il codec che arriva ora, AV2, è classico.
«MLVC riduce la banda dell'88%»	Rispetto a H.264 a 360p, sul set di test della stessa Microsoft. Rispetto a HEVC hardware è il 75,5% a 360p e il 65,4% a 540p, in chiamate in cui Microsoft controlla entrambe le estremità.
«Le texture neurali riducono la memoria video dell'85–96%»	Nelle scene dimostrative. Il gioco pubblicato ha guadagnato circa il 30% di compressione effettiva e ha usato le texture neurali solo su asset selezionati a causa del costo in esecuzione.
«JPEG AI è uno standard ISO, quindi fotocamere e browser lo adotteranno»	Non ho trovato supporto di default nei prodotti. Il formato che il web ha adottato nel 2026, JPEG XL in Chrome 155, è classico.
«La decodifica neurale è troppo costosa per i telefoni»	In parte vero. MLVC decodifica 540p30 entro metà di una NPU; il 4K60 con un modello di classe DCVC-RT richiederebbe circa 175 TOPS di calcolo sostenuto stima.
«I collegamenti satellitari e via drone sono il grande mercato del video neurale»	Tecnicamente l'uso più adatto, ma un mercato piccolo e sempre più soggetto a controlli all'esportazione su entrambe le sponde del Pacifico.
«Le grandi piattaforme prenderanno in licenza un codec neurale da una startup»	Hanno costruito il proprio (Netflix, Microsoft, Tencent) o hanno comprato la startup (Apple, Sony, InterDigital).

Un test semplice smonta gran parte delle presentazioni commerciali: 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.

CAPITOLO 12

12 INDICATORI
Cosa osservare, dal 2027 al 2030

Quando	Indicatore	Perché conta
Gennaio 2027	JVET valuta a Brisbane le risposte per H.267	Quante delle risposte migliori si basano su strumenti neurali, e a quale costo per pixel
Inizio 2027	I primi stream web AV2 di YouTube; decoder AV2 nei SoC annunciati	Il prossimo ciclo classico fissa i budget dei dispositivi fino al 2030 circa
Seconda metà del 2027	Il primo modello di test di H.267	Se gli strumenti neurali sono nel progetto di base o solo in esplorazione
2027	MLVC oltre le chiamate Teams peer-to-peer; altri fornitori che adottano il codice aperto	Il primo test di un codec neurale aperto fuori dall'azienda che lo possiede
2027–2028	Un primo prodotto che scrive o legge JPEG AI di default	Il primo formato neurale usato oltre i confini di un'azienda
2027–2028	DirectX Linear Algebra nei driver in commercio; texture neurali in più giochi e su mobile	Decodifica neurale indipendente dal produttore dentro la grafica
In corso	Budget delle NPU in telefoni e PC	A circa 100 TOPS sostenuti, un decoder 1080p30 di classe DCVC-RT occuperebbe circa un quinto della NPU anziché la metà analisi
Fine 2029	Versione 1 di H.267	L'inizio del ciclo del silicio per gli strumenti neurali nella riproduzione di massa

CAPITOLO 13

13 QUESTIONI APERTE Questioni aperte

Diverse questioni di questo rapporto non hanno ancora una risposta pubblica, e sono quelle che trovo più interessanti sul piano tecnico.

Esattezza su larga scala. Quanto può avvicinarsi un decoder neurale a un'aritmetica esatta al bit, solo su interi, su NPU di produttori diversi prima di rinunciare a gran parte del suo vantaggio di compressione? Il compromesso di sei punti di MLVC è un dato; uno standard ha bisogno di una risposta generale.

La memoria prima dei MAC. Quanto costa la post-elaborazione neurale in traffico DRAM per pixel quando gira al volo tra un decoder hardware e il display, e quanto si può ridurre il suo working set?

Il prezzo di uno strumento neurale in H.267. A quanti MAC e nanojoule per pixel un filtro in-loop neurale smette di ripagare i bit che fa risparmiare in un decoder 4K60?

Circuiti chiusi oltre le chiamate. Quali altri sistemi chiusi, dall'archiviazione di foto nel cloud e dallo streaming di giochi fino ai robot e ai collegamenti di contribuzione privati, possono adottare codec neurali end-to-end prima dell'arrivo di H.267?

Decoder che generano. Quando un decoder genera dettagli invece di ricostruirli, come fanno già i VAE video, come andrebbe misurata e specificata la fedeltà?

Sono temi che possiamo esplorare insieme nei prossimi anni, attraverso lo scambio tecnico e la collaborazione tecnica: i@tanshuai.com.

Limiti di questo rapporto

Questa è una ricerca documentale basata su documenti pubblici fino al 12 novembre 2026. Per questa edizione non ho condotto interviste primarie né misure indipendenti. Diversi numeri chiave sono dichiarazioni di produttori o di autori sui propri set di test, come i risparmi di bitrate di MLVC e i rapporti di compressione delle texture neurali; sono indicati come tali. Le mie stime di calcolo per DCVC-RT sommano i valori per modulo pubblicati dai suoi autori e ignorano le perdite di utilizzo, quindi il carico reale della NPU sarà più alto. Il confronto energetico si basa su pochi progetti pubblicati su nodi di processo diversi; i dati dei decoder VVC e AV2 di produzione sono proprietari. Le tempistiche degli standard sono gli obiettivi dichiarati dai comitati stessi. Le date di H.267 successive al 2029 sono mie stime basate sulle storie di AV1 e VVC.

L'autore

Tan Shuai è un tecnologo e fondatore con più di vent'anni di lavoro tra chip, hardware e software di sistema, e 20 brevetti concessi. Ha alle spalle molti anni di R&S e di gestione nel settore dei chip fabless, tra cui la progettazione full-custom di datapath ASIC e progetti di chip multimediali per fotocamere come acceleratori di codec video, ed è questa la lente di questo rapporto: energia per pixel, traffico di memoria e ciò che uno standard consente a un decoder di fissare nel silicio. Il suo lavoro lo ha portato nella Silicon Valley, a New York, a Chicago, a Taiwan, nella Greater Bay Area (Shenzhen e Hong Kong), in Vietnam e in Malesia. Oggi lavora come ricercatore e sviluppatore nel campo dell'IA.

Per discussioni tecniche: i@tanshuai.com · tanshuai.com

Note e avvertenze

Questo rapporto è una ricerca documentale indipendente basata su documenti pubblici disponibili fino al 12 novembre 2026. È fornito a solo scopo informativo e non costituisce consulenza in materia di investimenti, legale o di ingegneria. Prodotti, tempistiche degli standard, tariffe dei pool di brevetti e regole sull'esportazione cambiano; è bene verificare sulla fonte primaria qualsiasi elemento su cui si fa affidamento. I valori contrassegnati come **dato del produttore**, **stima** o **analisi** vanno letti come tali. I nomi di aziende e prodotti appartengono ai rispettivi proprietari e sono usati solo a scopo identificativo; nessuna azienda ha revisionato o sponsorizzato questo rapporto.

© 2026 Tan Shuai. È consentito condividere questo rapporto senza modifiche, citandone l'autore.

Fonti

Tutte consultate tra il 3 e il 12 novembre 2026. Raggruppate per capitolo.

Traffico, dispositivi e ambito

- Ericsson, *Mobility Report*, giugno 2026 (video circa il 75% del traffico dati mobile alla fine del 2025): https://www.ericsson.com/en/reports-and-papers/mobility-report/reports/june-2026 ; sintesi: 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, consegne mondiali di smartphone, 4º trimestre e intero 2025 (1,26 miliardi), 13 gennaio 2026: https://www.idc.com/resource-center/press-releases/q425mobilephonetop5/
- TrendForce, previsione delle consegne di TV per il 2025 (195,71 milioni): 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, previsione TV per il 2025 (208,7 milioni): https://iconnect007.com/article/145954/omdia-forecasts-flat-tv-shipments-in-2025-but-mini-led-poised-for-explosive-growth-at-829/145951/ein

Riproduzione e streaming

- Netflix Technology Blog, «AV1 — Now Powering 30% of Netflix Streaming», dicembre 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», novembre 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 maggio 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» (quota AV1 del catalogo di 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» (esercizio commerciale del DTV+, giugno 2026): https://mpegh.com/brazil/ ; CartaCapital sui ricevitori nei televisori: https://www.cartacapital.com.br/?p=628337 ; Olhar Digital su VVC nella 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 ottobre 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 ottobre 2025: https://atscaleconference.com/how-meta-deployed-super-resolution-at-scale-to-transform-video-quality/
- Samsung, caratteristiche principali dei TV 8K (NQ8 AI Gen3, 768 reti neurali): https://www.samsung.com/us/tvs/8k-tv/highlights/

Fotografia e immagini

- 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 e Visionary.ai, ISP completo basato sull'IA, 5 gennaio 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/ ; intervista di PetaPixel: https://petapixel.com/2026/01/18/a-look-at-the-worlds-first-full-ai-based-image-signal-processor
- Comitato JPEG, comunicato stampa della 103ª riunione (guadagni di JPEG AI, complessità, implementazioni su smartphone): https://jpeg.org/items/20240507_press.html
- G. J. Sullivan, introduzione a ISO/IEC JTC 1/SC 29, aprile 2026 (stato di JPEG AI; dati di ECM e 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/
- Gruppo di lavoro IEEE 1857.11 (codifica di immagini basata su reti neurali): https://sagroups.ieee.org/fvc
- Chrome 155 e JPEG XL, 6 ottobre 2026: https://xenospectrum.com/en/chrome-155-jpeg-xl-rust/

Chiamate e riunioni

- Microsoft, «Satin: Microsoft's latest AI-powered audio codec for real-time communications», 17 febbraio 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 luglio 2026: https://techcommunity.microsoft.com/blog/linuxandopensourceblog/announcing-the-open-source-release-of-ml-video-codec-mlvc/4539875 ; codice: 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 ; analisi di iReplay: https://ireplay.tv/blog/mlvc-microsoft-neural-video-codec-open-source-streaming/

- Note di rilascio di 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 giugno 2024: https://engineering.fb.com/2024/06/13/web/mlow-metas-low-bitrate-audio-codec/
- Microsoft, PC Copilot+ (NPU da oltre 40 TOPS): https://www.microsoft.com/en-us/windows/copilot-plus-pcs ; Apple, annuncio dell'M4 (Neural Engine da 38 TOPS): https://www.apple.com/newsroom/2024/05/apple-introduces-m4-chip/

IA generativa

- Kyutai, scheda del modello Mimi (12,5 Hz, 1,1 kbps): https://huggingface.co/kyutai/mimi
- Lightricks, «LTX-Video: Realtime Video Latent Diffusion» (VAE video 1:192): https://arxiv.org/abs/2501.00103
- Comunicato stampa della 153ª riunione MPEG (Video Coding for Machines allo stadio DIS): https://www.mpeg.org/wp-content/uploads/2026/03/press_release_153.pdf

Videogiochi e grafica

- NVIDIA, adozione di DLSS 4: lancio e Computex 2025 (75 e 125+): https://www.nvidia.com/en-us/geforce/news/125-dlss-4-multi-frame-gen-games-more-announced-computex-2025/ ; Gamescom 2025 (175+): https://www.nvidia.com/en-us/geforce/news/gamescom-2025-dlss-4-full-ray-tracing-game-announcements/ ; CES 2026 (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», febbraio 2026: https://www.ubisoft.com/en-us/studio/laforge/news/415slmB3ZGzv8d2LhO9QT/shipping-neural-texture-compression-in-assassins-creed-mirage
- TweakTown sulla demo di texture neurali di NVIDIA al GTC 2026 (da 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 DirectX di Microsoft, Shader Model 6.9 in versione definitiva, 26 febbraio 2026: https://devblogs.microsoft.com/directx/shader-model-6-9-retail-and-more/ ; sul cooperative vector: https://devblogs.microsoft.com/directx/shader-model-6-9-and-the-future-of-cooperative-vector/ ; piani per la GDC 2026: https://windowsreport.com/microsoft-showcases-neural-rendering-and-dxr-2-0-plans-for-directx-at-gdc-2026/

Collegamenti specializzati

- Global Trade Alert sull'Annuncio n. 49 del 2026 della Cina: 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 (soglie per le piattaforme terrestri senza equipaggio): 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 aggiunge i droni prodotti all'estero e i loro componenti alla Covered List, 22 dicembre 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 ; pagina del chip al 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 ; materiale supplementare (complessità per modulo): https://openaccess.thecvf.com/content/CVPR2025/supplemental/Jia_Towards_Practical_Real-Time_CVPR_2025_supplemental.pdf
- NNVC e la libreria di inferenza su interi SADL: https://arxiv.org/abs/2309.05846
- Tencent Canghai V2, 27 maggio 2026 (Shanghai Securities News tramite 10jqka): https://stock.10jqka.com.cn/20260527/c677016020.shtml ; sintesi delle funzionalità: https://www.besthub.dev/articles/tencent-s-canghai-v2-chip-enters-mass-production-what-the-new-capabilities-mean-for-video-encoding-bcfc18dbfa29
- IP di decoder AV2 di Chips&Media, 12 maggio 2026: https://www.design-reuse.com/news/202530501-chips-media-completes-development-of-next-gen-av2-hw-decoder-ip/ ; licenza AV2, 17 giugno 2026: https://www.edge-ai-vision.com/?p=58270
- Allegro DVT, IP di decoder AV2, 16 giugno 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 giugno 2026: https://www.businesswirechina.com/en/news/63176.html

Standard e brevetti

- Comunicato stampa della 155ª riunione MPEG (richiesta congiunta di proposte oltre VVC), 5 agosto 2026: https://www.mpeg.org/wp-content/uploads/2026/08/ag3w254SN26474Press-Release-MPEG155.pdf
- Comunicato stampa della 152ª riunione MPEG (risultati della richiesta di evidenze; edizioni di VVC e VSEI), 30 ottobre 2025: https://www.mpeg.org/wp-content/uploads/2025/11/Press-Release-MPEG152.pdf ; rubrica MPEG di 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 brevettuale oltre VVC (H.267): https://greyb.com/blog/exclusive-video-codec-landscape/ ; sulle licenze di AV2 e la vendita del portafoglio di Alibaba: https://greyb.com/blog/av2-licensing/
- MLex, «Fragmented codec licensing exposes need for clearer rules for Chinese tech majors», 27 marzo 2026: https://www.mlex.com/mlex/articles/2458402/fragmented-codec-licensing-exposes-need-for-clearer-rules-for-chinese-tech-majors
- Access Advance, tariffe di royalty di MCBA e VVC Advance al 1º luglio 2026: https://accessadvance.com/wp-content/uploads/2025/07/MCBA-Program-Overview-September-2026.pdf

Denaro e aziende

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

COMPRESSIONE APPRESA · RIPRODUZIONE × SILICIO

La codifica neurale nel 2026

Si usa attorno al codec, non al suo posto

Ogni numero chiave di questo rapporto riporta la sua fonte e la sua data. Dove l'autore ricava un numero, questo è indicato come stima; dove lo dichiara solo chi vende il prodotto, è indicato come dato del produttore.

Tan Shuai

Tecnologo e fondatore · Ricercatore e sviluppatore di IA

tanshuai.com · i@tanshuai.com

EDIZIONE PUBBLICA · Dati aggiornati al 12 novembre 2026