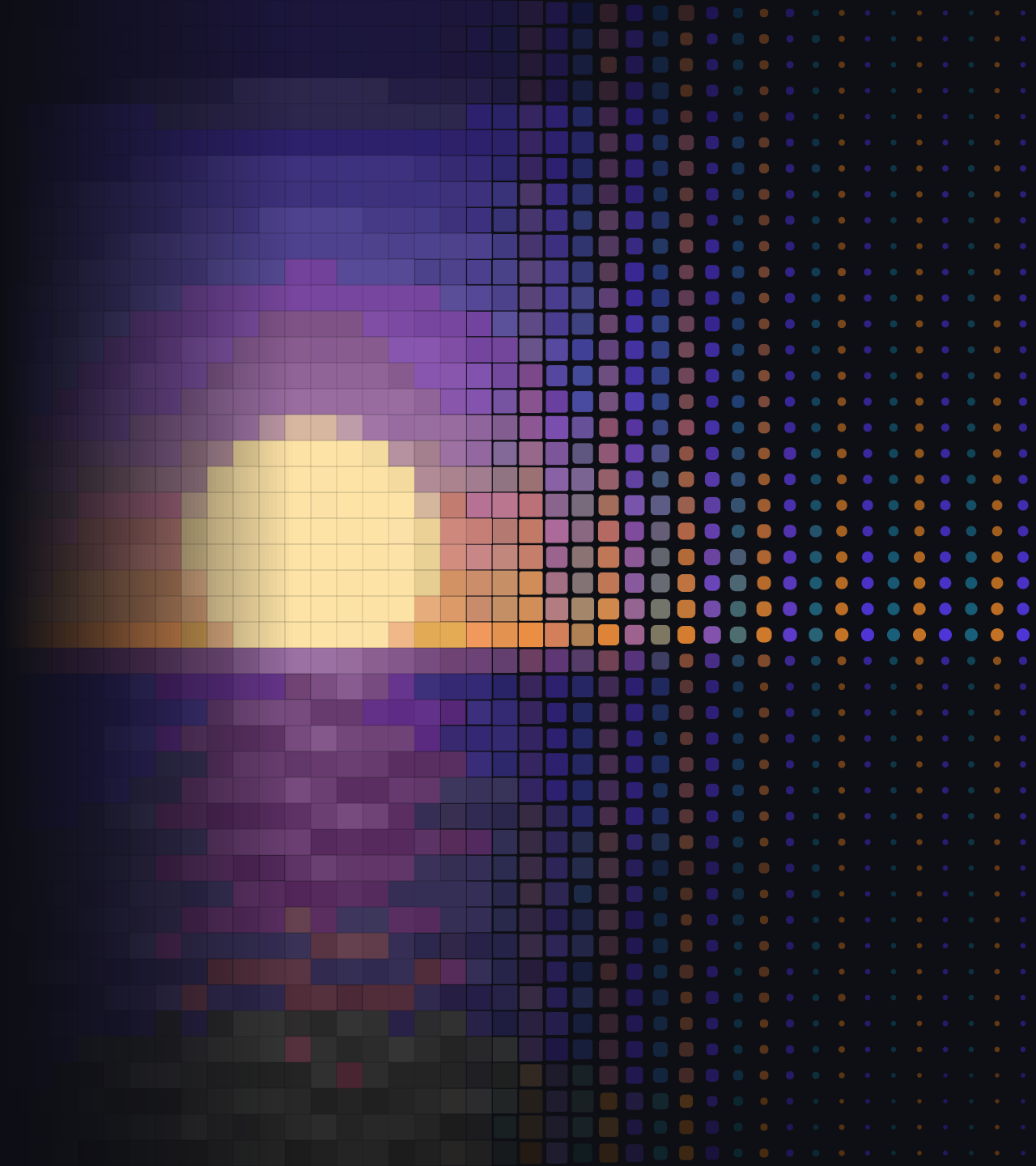
Forschungsnotiz · November 2026

Neuronale Codierung 2026

Rund um den Codec, nicht an seiner Stelle

Wiedergabe · Fotografie · Telefonie · Generative KI · Spiele · Silizium

Tan Shuai
tanshuai.com · i@tanshuai.com



DAS FAZIT

Bereits im Mainstream, aber rund um den Codec

01 Neuronale Netze sind an beiden Enden der Standard-Codecs etabliert

Kamera-ISPs, Downscaling vor dem Encoder, Super-Resolution bei der Wiedergabe. Dazwischen bleibt der Bitstrom AV1, HEVC, AVC oder VVC, decodiert von festverdrahteter Hardware.

02 Neuronale End-to-End-Codecs gibt es nur in geschlossenen Systemen

Anrufe, Tokenizer generativer KI und ein AAA-Spiel: Ein Unternehmen kontrolliert beide Enden und kann auf einen Standard-Codec ausweichen.

03 Neuronale Bitströme im Massenmarkt warten auf H.267

Version 1 ist für Ende 2029 geplant; erste Chips um 2031, breiter Einsatz Mitte der 2030er-Jahre Schätzung.

≈75 %

des mobilen Datenverkehrs ist Video (Ende 2025)

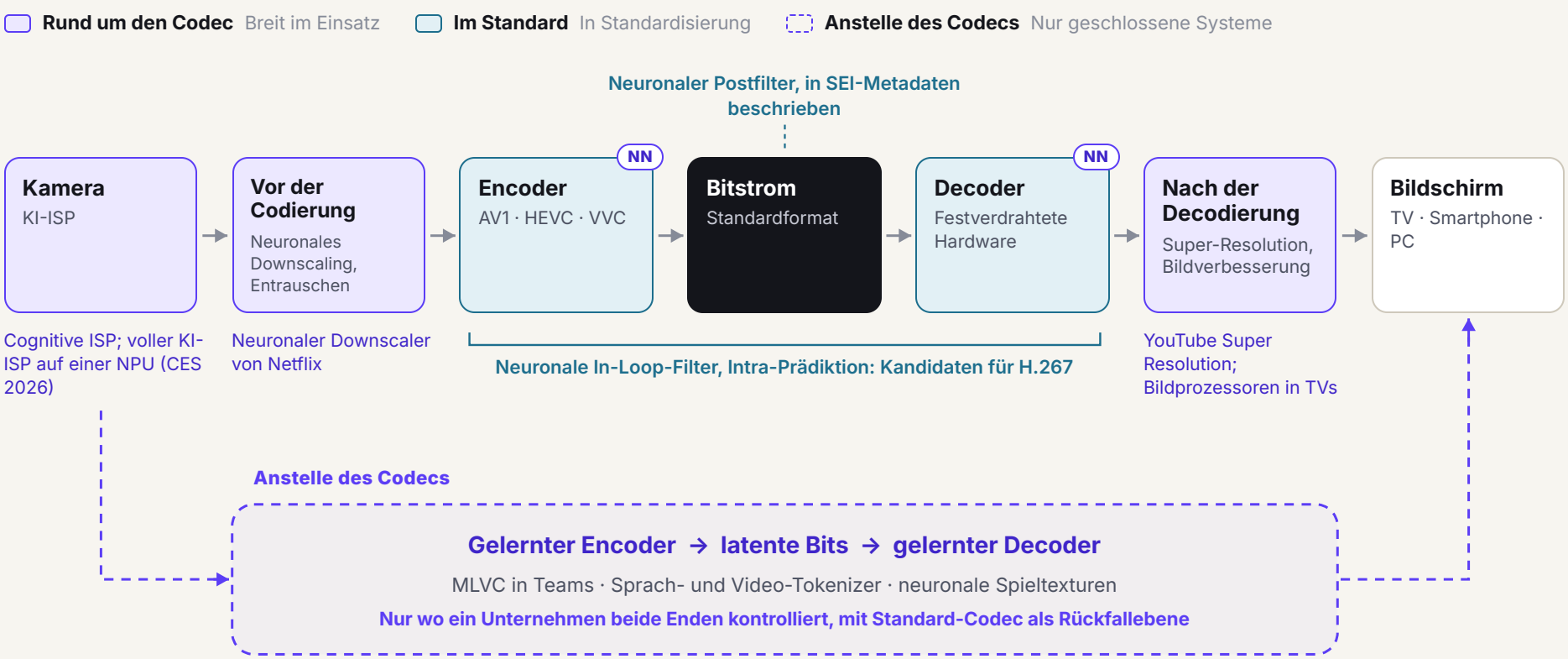
1,26 Mrd.

2025 ausgelieferte Smartphones, alle mit Videodecodierung in festverdrahteter Hardware

≈12×

Energie pro Pixel gegenüber einem Hardware-HEVC-Decoder: das einzige veröffentlichte neuronale Decoder-Chipdesign

Drei Wege neuronaler Netze in den Codec



> Rund um den Codec braucht es niemandes Erlaubnis. Im Standard braucht es einen Standard. Anstelle des Codecs braucht es die Kontrolle über beide Enden.

Wo neuronale Codierung 2026 im Einsatz ist

	Rund um einen Standard-Codec Vorverarbeitung, Bildverbesserung, Upscaling	Neuronaler End-to-End-Codec gelernter Encoder und Decoder	Im Standard neuronale Werkzeuge in Standard-Bitstrom oder API
01 **Wiedergabe und Streaming**	BREIT IM EINSATZ Netflix-Downscaler; Super-Resolution bei YouTube und Meta; KI-Upscaler in TVs	NICHT IM EINSATZ Kein Streamingdienst sendet ein neuronales Videoformat; TVs und Smartphones decodieren Standard-Codecs in Hardware	IN STANDARDISIERUNG H.267 Version 1 kommt Ende 2029; neuronale Werkzeuge gehören zu den Kandidaten
02 **Fotografie und Bilder**	BREIT IM EINSATZ KI-ISPs in Flaggschiff-Smartphones; voll neuronaler ISP als IP, gezeigt auf der CES 2026	DEMOS, PILOTPROJEKTE JPEG AI ist ISO-Standard: Smartphone-Demos, kein Produkt	NICHT IM EINSATZ Das Format, das an Boden gewinnt, ist klassisch: JPEG XL in Chrome 155 standardmäßig aktiv
03 **Telefonie und Konferenzen**	BREIT IM EINSATZ Empfängerseitige Super-Resolution und neuronale Geräuschunterdrückung	BREIT IM EINSATZ Neuronales Audio in Teams-Zweiergesprächen seit 2021; MLVC-Video wird in Teams eingeführt	BEGRENZT IM EINSATZ Opus 1.5 ergänzt optionale, abwärtskompatible ML-Werkzeuge
04 **Generative KI**	—	BREIT IM EINSATZ Jedes Video- und Voice-Modell nutzt einen gelernten Tokenizer, z. B. einen Video-VAE mit 1:192	—
05 **Spiele und Grafik**	BREIT IM EINSATZ Neuronales Upscaling und Frame Generation: allein DLSS 4 in über 250 Spielen und Apps	BEGRENZT IM EINSATZ Neuronale Texturen in einem AAA-Spiel ausgeliefert; NVIDIAs SDK ist in der Beta	IN STANDARDISIERUNG DirectX erhält Matrixoperationen für Shader (Vorschau)
06 **Spezialverbindungen: Satellit, Roboter**	DEMOS, PILOTPROJEKTE Serverseitige Restaurierung von Feeds mit niedriger Bitrate	DEMOS, PILOTPROJEKTE Verlusttolerante neuronale Codecs in Forschung und Pilotprojekten	—

Die Zeilen sind nach der Reichweite jedes Bereichs geordnet.

Neuronal an beiden Enden, Standard in der Mitte

- **Am Encoder**

  Der neuronale Downscaler von Netflix läuft seit 2022 ohne Änderung an irgendeinem Gerät; in den Tests von Netflix bevorzugten ihn etwa 77 % der Zuschauer Anbieterangabe.

- **Am Bildschirm**

  YouTube Super Resolution auf Fernsehern seit Oktober 2025; Meta betreibt Super-Resolution auf seinen Servern und auf Smartphones.

- **Im Fernseher**

  Samsungs Prozessor NQ8 AI Gen3 wirbt mit 768 neuronalen Netzen für das Upscaling.

DER BITSTROM BLEIBT STANDARDKONFORM

30 %

des Netflix-Streamings in AV1 (Dezember 2025)

über 90 %

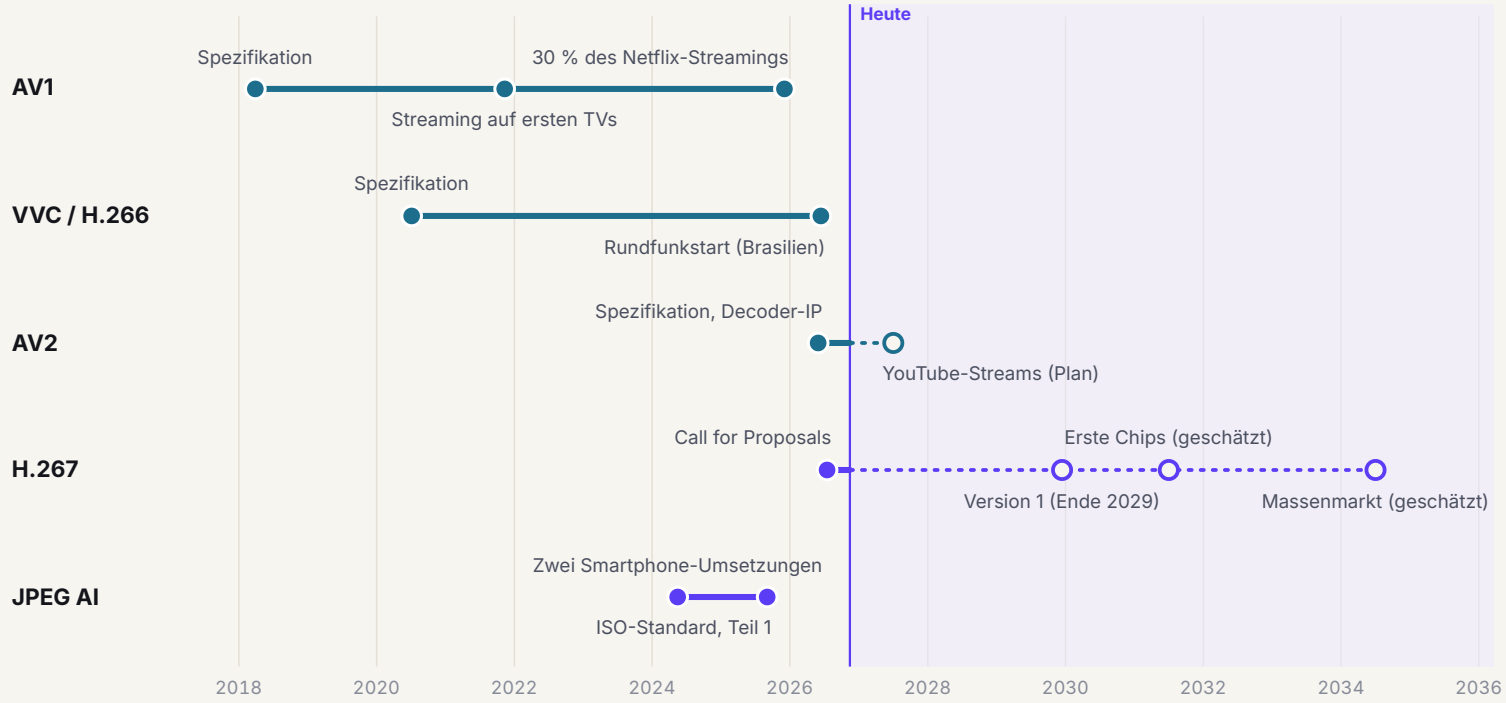
des YouTube-Katalogs, gewichtet nach Wiedergabezeit, in AV1 verfügbar (März 2026)

3 Wochen

von AV2 Version 1.0.0 (28. Mai 2026) bis zu Decoder-IP-Kunden bei drei Anbietern

ZEITPLAN

Vom fertigen Standard bis zum Massenmarkt vergeht der Großteil eines Jahrzehnts



Von der Spezifikation (2018) bis zu 30 % des Netflix-Streamings (Dezember 2025) brauchte AV1 fast acht Jahre. Hohle Marken sind Ziele oder meine Schätzungen.

H.267 entscheidet, ob neuronale Werkzeuge jeden Bildschirm erreichen

Oktober 2025
Call for Evidence: fünf Antworten von 16 Firmen und Laboren; neuronale Gewinne am klarsten bei 8K HDR

Juli 2026
Call for Proposals in Genf veröffentlicht

Januar 2027
Einreichungen in Brisbane ausgewertet

Zweite Hälfte 2027
Erstes Testmodell

Ende 2029
Version 1 des Standards

≈28 %
spart der klassische ECM-Zweig gegenüber VVC

6–14 %
sparen neuronale In-Loop-Filter bei 10 bis über 22 Tausend MACs pro Pixel

SEI
VVC-Streams können schon heute einen neuronalen Postfilter für den Decoder signalisieren

Meine Lesart: ein klassischer Kern plus einige neuronale Werkzeuge, wo die gesparten Bits den Decodieraufwand rechtfertigen Analyse.

Neuronal in der Kamera, klassisch im Web

- **Kamera**

  Qualcomms Cognitive ISP segmentiert seit 2023 in Smartphones in Echtzeit; auf der CES 2026 war ein vollständiger KI-ISP zu sehen, der auf einer NPU läuft.

- **JPEG AI**

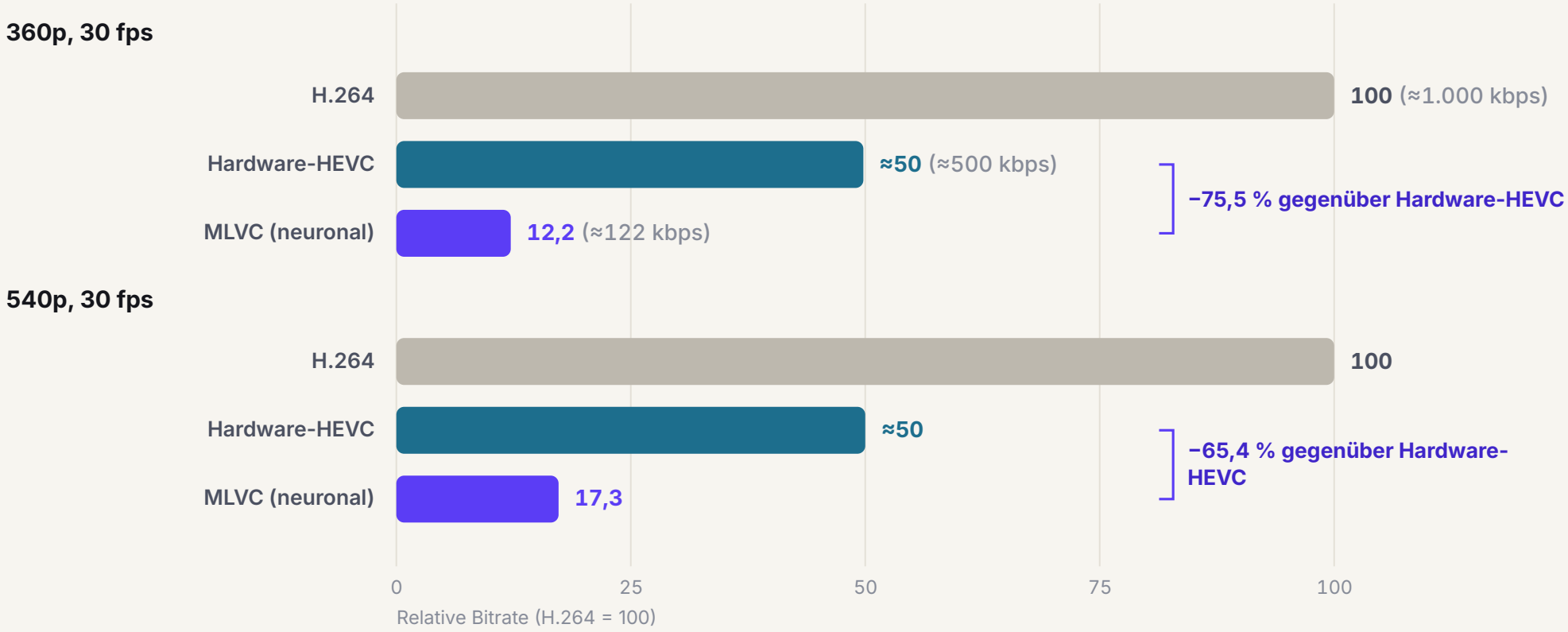
  2025 veröffentlicht: 12,5–27,9 % gegenüber VVC Intra bei 7 bis 216 Tausend MACs pro Pixel. Ich fand kein Produkt, das JPEG AI standardmäßig nutzt.

- **Web**

  Chrome 155 hat JPEG XL am 6. Oktober 2026 standardmäßig aktiviert, mit einem speichersicheren Decoder in Rust.

Formate, die überall lesbar sein müssen, bleiben klassisch; neuronale Codierung gewinnt in der Pipeline eines Unternehmens.

Bei Anrufen sind neuronale End-to-End-Codecs im Einsatz



Satin läuft seit 2021 in allen Zweiergesprächen in Teams und Skype. MLVC, seit dem 24. Juli 2026 Open Source unter MIT-Lizenz, wird in Peer-to-Peer-Anrufen in Teams mit bis zu 540p30 eingeführt und belegt höchstens die Hälfte der NPU **Anbieterangabe**.

Nicht jede Plattform wählte neuronale Codecs, und die Gründe sind lehrreich

IM STANDARD, OPTIONAL

Opus 1.5 und 1.6

Neuronale Verbesserung, Verschleierung von Paketverlusten, Deep Redundancy und Bandbreitenerweiterung, alle standardmäßig aus und voll kompatibel mit dem Standard von 2012.

KLASSISCH, BEWUSST GEWÄHLT

Meta MLow (2024)

Über 20 % der Meta-Anrufe laufen auf ARMv7-Smartphones, täglich Dutzende Millionen WhatsApp-Anrufe auf über zehn Jahre alten Geräten. MLow bietet bei 6 kbps etwa doppelte Opus-Qualität mit 10 % weniger Rechenaufwand.

Das Komplexitätsbudget des Gerätebestands entscheidet. Neuronale Codecs kommen zuerst, wo Geräte neu sind und ein Anbieter beide Enden kontrolliert.

Der größte Einsatz, den niemand Codec nennt

- **Überall gelernte Codecs**

  Jedes Voice- und Videomodell liest und schreibt die Tokens oder latenten Werte eines gelernten Codecs.

- **Keine Distributionsformate**

  Jedes Modell hat seinen eigenen Latentraum: keine Ratenregelung, kein Konformitätstest, kein zweiter Decoder.

- **Ein Kostenposten**

  Wer generatives Video anbietet, senkt mit einer stärker komprimierenden latenten Darstellung die Generierungskosten überproportional Analyse.

ZWEI BEISPIELE

12,5 Hz · 1,1 kbps

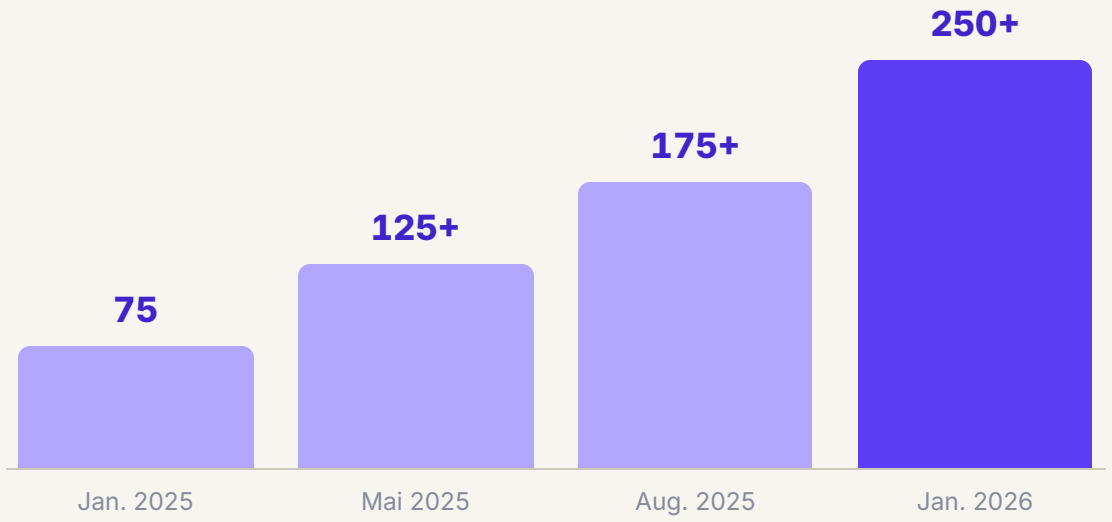
Sprach-Tokens von Kyutais Mimi, unter jedem Telefonie-Codec

1:192

Video-VAE von LTX-Video: ein latentes Token für 32×32 Pixel über 8 Frames

Neuronale Rekonstruktion ist etabliert, neuronale Assets stehen am Anfang

DLSS 4 Multi Frame Generation: Spiele und Apps



Neuronale Texturkompression

NVIDIA-Demoszene, GTC 2026
Texturspeicher 6,5 GB → 0,97 GB

85 % weniger Speicher

Assassin's Creed Mirage, ausgeliefert
Nur ausgewählte Assets; gegenüber Blockformaten

≈30 % bessere Kompression

Unterschiedliche Maße: nicht direkt vergleichbar

DirectX Linear Algebra ging im April 2026 in die öffentliche Vorschau: ein Codepfad für neuronale Decodierung in Shadern auf den GPUs aller Hersteller.

Satelliten, Drohnen und Roboter: die beste Eignung, ein kleiner Markt

- **Warum es passt**

  Ein Betreiber kontrolliert beide Enden, Bandbreite ist knapp, die Inhalte sind eng genug für das Training.

- **China**

  Bekanntmachung Nr. 49 von 2026 (gültig ab 18. November) überarbeitet die Drohnen-Exportkontrollen; ein Eintrag nutzt semantische Bildkompression als Schwelle.

- **Vereinigte Staaten**

  Die FCC setzte am 22. Dezember 2025 im Ausland gefertigte Drohnen und deren kritische Komponenten auf ihre Covered List.

> **Eine solide Nische für einen Spezialisten mit Kapazitäten für Export-Compliance, kein Massenmarkt.**

Eine Größenordnung mehr Energie pro Pixel

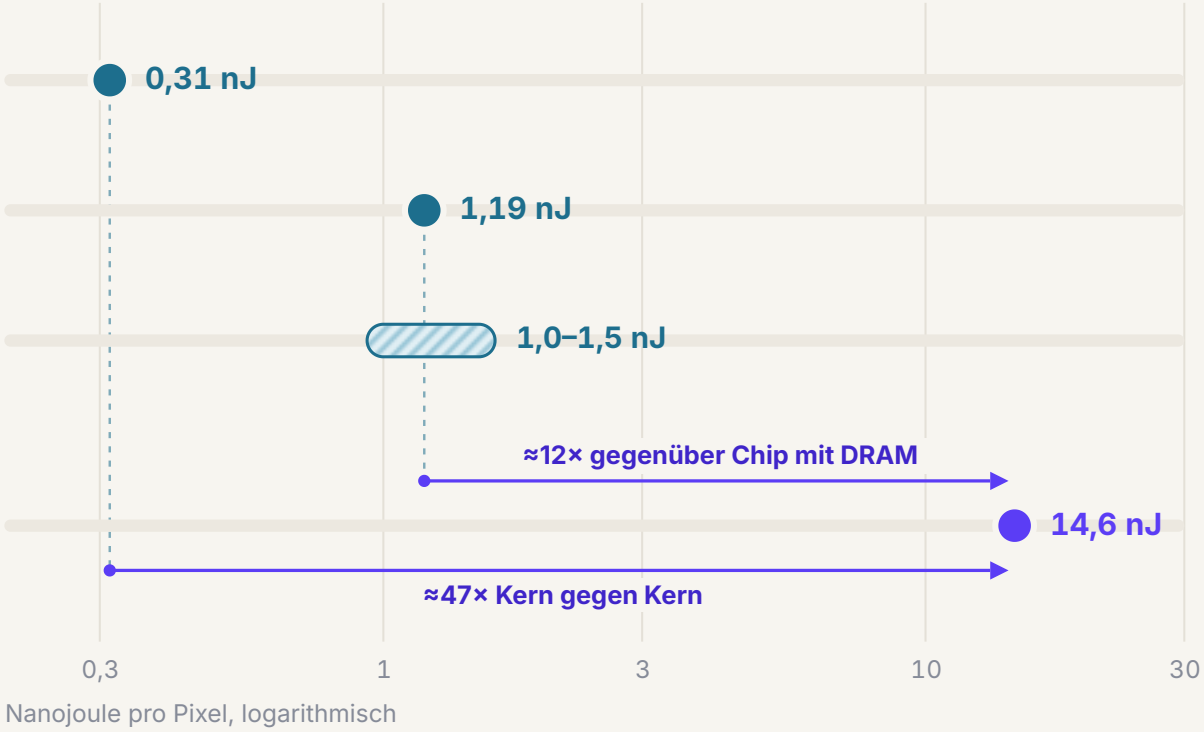
HEVC-Decoderchip, Kern
40 nm, 4K30, gemessen

HEVC-Decoderchip mit DRAM
40 nm, 4K30, Kern gemessen, DRAM simuliert

VVC-Decoder
28 nm, 4K60, Projektion

Neuronaler Videodecoder (NVCA)
28 nm, 1080p25, nur Synthese; ohne Gewichte und DRAM

Bei 1080p30 sind das etwa 74 mW gegenüber 908 mW.



Bei 1080p30: etwa 74 mW für den HEVC-Chip mit DRAM gegenüber rund 900 mW für NVCA. Älterer Fertigungsprozess beim HEVC-Chip, Gewichte und DRAM bei NVCA nicht mitgezählt: Der Vergleich begünstigt das neuronale Design.

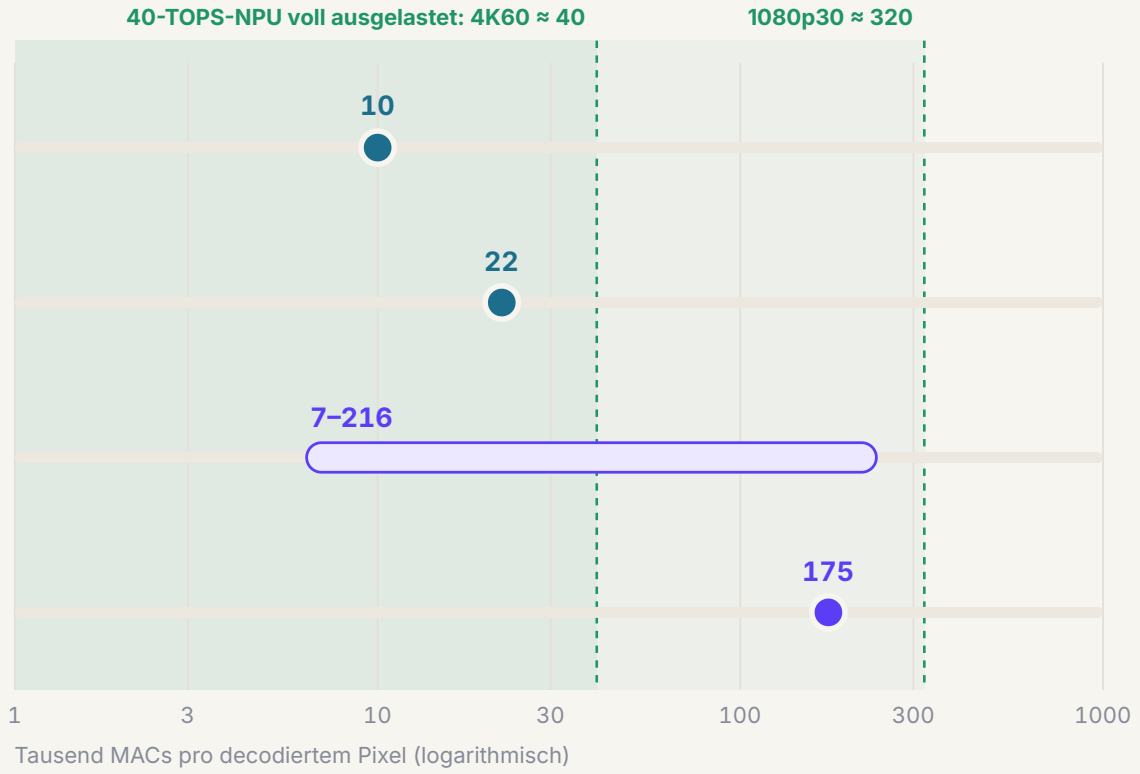
Was ein neuronaler Decoder von einer NPU verlangt

Neuronaler In-Loop-Filter, sehr geringe Komplexität
JVET-Exploration · ≈6 % Bitersparnis
1,2 TOPS bei 1080p30 · 10 bei 4K60

Neuronaler In-Loop-Filter, geringe Komplexität
JVET-Exploration · ≈8 %
2,7 TOPS bei 1080p30 · 22 bei 4K60

JPEG-AI-Decoder, Basis- bis High-Variante
12,5–27,9 % gegenüber VVC Intra

DCVC-RT, ein ganzer neuronaler Videodecoder
21 % gegenüber VVC-Referenz, Testsetup des Papers
22 TOPS bei 1080p30 · 174 bei 4K60



DCVC-RT braucht etwa 175 Tausend MACs pro Pixel: etwa 22 TOPS bei 1080p30 und 174 bei 4K60, bei voller Auslastung **Schätzung**. Eine voll ausgelastete 40-TOPS-NPU schafft bei 4K60 etwa 40 Tausend.

Speicherverkehr und Bitgenauigkeit sind die harten Grenzen

DRAM DOMINIERT

Speicher vor MACs

Im HEVC-Chip des MIT entfallen etwa drei Viertel der Systemenergie auf DRAM: 0,88 von 1,19 nJ pro Pixel. Neuronale Decoder tragen große Feature-Maps von Frame zu Frame mit; laut den DCVC-RT-Autoren bremste Speicher-I/O, nicht die Arithmetik.

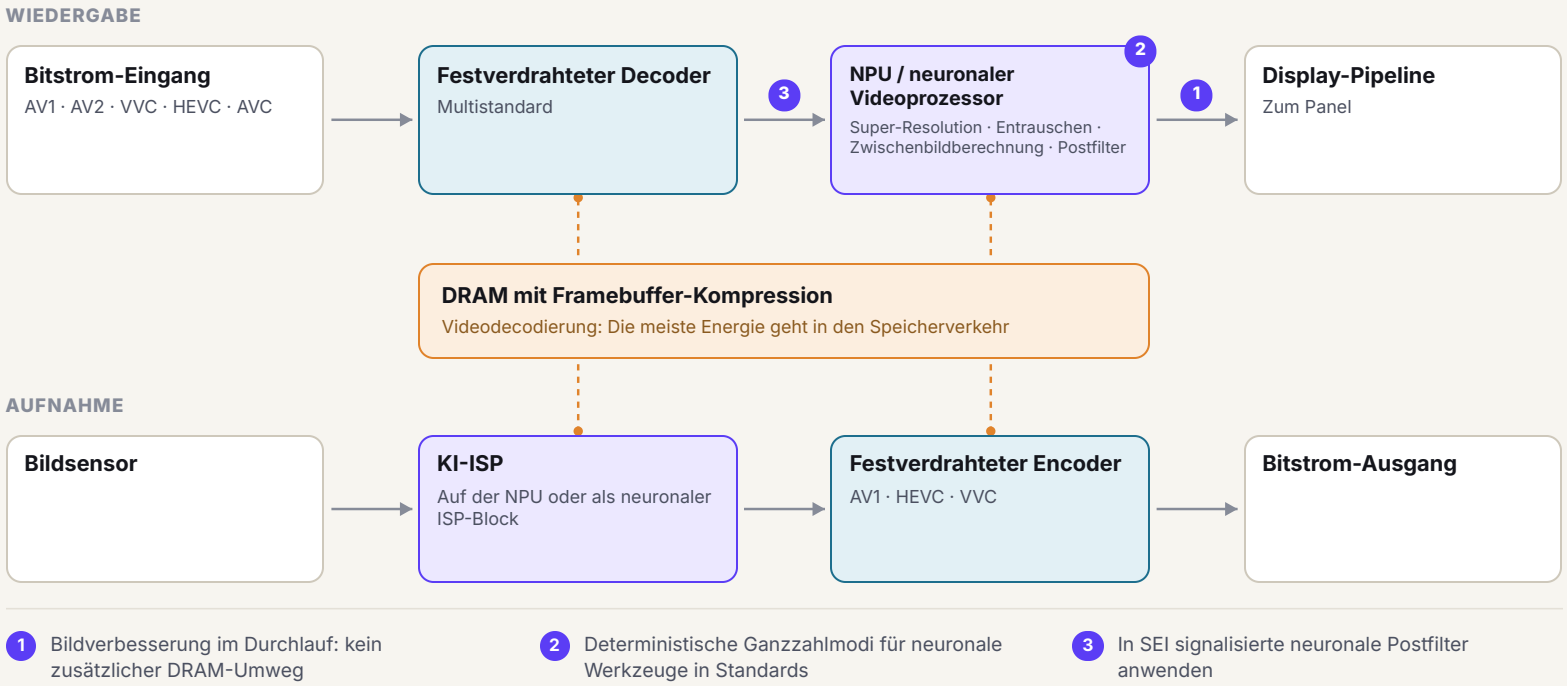
EXAKTHEIT

Jeder Decoder muss dieselben Bits berechnen

Die Entropiedecodierung scheitert, wenn zwei Geräte eine Wahrscheinlichkeit verschieden berechnen. Lösungen: Ganzzahlmodelle (DCVC-RT), Hyperprior mit gemeinsamer Skala (MLVC), 16-Bit-Ganzzahl-Inferenz (SADL von JVET), Konformitätspunkte (JPEG AI).

Für ein Chipteam ist das wieder die Disziplin der Ganzzahltransformationen von HEVC: Der Standard muss die Arithmetik exakt festlegen.

Die Aufteilung in Serienchips: festverdrahteter Codec plus NPU



> Canghai von Tencent: mehr als 100.000 Chips der ersten Generation im Einsatz; die zweite Generation bringt H.266 und Super-Resolution auf dem Chip. Chips&Media und Allegro DVT verkaufen Codec-IP und neuronale Video-IP nebeneinander.

Was ich bauen würde, nach Rolle

SoC- und TV-Chip-Teams

- Festverdrahtete Decoder für alle Pflichtstandards, AV2 jetzt
- NPU, ausgelegt auf Bildverbesserung in Displayauflösung, ohne zusätzlichen DRAM-Umweg
- SEI für neuronale Postfilter und deterministische Ganzzahlmodi
- H.267 als Designstart 2029 oder später

IP-Anbieter

- AV2- und VVC-Decoder jetzt
- Postfilter-Beschleuniger und Framebuffer-Kompression
- Konformitätswerkzeuge für neuronale Verarbeitung

Videoplattformen

- Neuronale Vorverarbeitung am Encoder zahlt sich zuerst aus
- Clientseitige Super-Resolution, wo die Geräte es erlauben
- Neuronale End-to-End-Codecs nur in eigenen geschlossenen Systemen

Wo neuronale Werkzeuge lizenzierbar werden

Standard	Neuronaler Inhalt	Stand November 2026
H.267 (Arbeitsname)	Offen: Klassische (ECM) und neuronale Werkzeuge sind beide Kandidaten	Auswertung der Einreichungen Januar 2027; Version 1 Ende 2029
VVC und HEVC mit VSEI	Neuronaler Postfilter in SEI beschrieben; für Decoder optional	In Kraft; im Oktober 2025 erweitert
AV2	Klassisch, lizenzgebührenfrei	Version 1.0.0 am 28. Mai 2026; YouTube-Webstreams für Anfang 2027 geplant
JPEG AI	Gelernter End-to-End-Bildcodec	Teil 1 veröffentlicht 2025; keine standardmäßige Produktunterstützung gefunden
IEEE 1857.11	Neuronale Bildcodierung, ausgelegt für NPUs	Veröffentlicht im Dezember 2024
Video Coding for Machines	Codierung für maschinelle Analyse statt für Betrachter	Internationaler Standardentwurf Anfang 2026
Opus-ML-Erweiterungen	Optional und abwärtskompatibel	In libopus 1.5 und 1.6

645

Patentfamilien von ByteDance jenseits von VVC (GreyB, 2025)

0,50 $ · 1,00–1,50 $

VVC-Pool-Tarif pro Smartphone und pro 4K-TV, Region 1

0 $

Lizenzgebühren der AOMedia-Mitglieder für AV1 und AV2

WER VERDIENT

Das Geld folgt den Standards und den Plattformen

VERDIENEN JETZT

Plattformen

Behalten die Ersparnis

Neuronaler Downscaler von Netflix; Super-Resolution bei YouTube und Meta; Microsoft hat MLVC unter MIT-Lizenz offengelegt

VERDIENEN JETZT

Geräte- und Chiphersteller

Verkaufen Differenzierung

Kamera-ISPs; TV-Upscaling-Prozessoren; Tencents hauseigene Canghai-Codec-Chips

VERDIENEN JETZT

IP-Anbieter

Lizenzieren jeden neuen Standard

AV2-Decoder-Kunden bei Chips&Media, Allegro DVT und VeriSilicon binnen drei Wochen nach der Spezifikation

VERDIENEN JETZT

GPU-Hersteller

Schaffen Plattformbindung

DLSS 4 in über 250 Spielen und Apps; SDK für neuronale Texturen in der Beta

PRO GERÄT

Patentinhaber und Pools

Kassieren pro Gerät

VVC: 0,50 $ pro Smartphone, 1,00–1,50 $ pro 4K-TV; H.267-Werkzeuge nach 2029

EXIT DURCH ÜBERNAHME

Start-ups für neuronale Codecs

Übernommen

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

> Die drei bekanntesten Start-ups für neuronale Codecs, WaveOne, iSIZE und Deep Render, wurden von Apple, Sony und InterDigital gekauft.

Sechs Geschäftsfelder und ihre Risiken

Geschäftsfeld	Wer kauft	Warum jetzt	Hauptrisiko
Neuronale Vorverarbeitung am Encoder	Streaming- und Nutzervideo-Plattformen, Sender	Keine Geräteänderung; bei Netflix erprobt	Große Plattformen bauen es selbst
IP und Tuning für Bildverbesserung am Display	TV-, Set-Top- und Smartphone-Chiphersteller	NPUs stecken schon in den Geräten	Kommoditisierung; hauseigene Teams
Neuronale Codecs für geschlossene Systeme	Anbieter für Unternehmenskommunikation, Industrie und Telepräsenz	MLVC beweist die Machbarkeit, der Code ist offen	Kontrolle über beide Enden nötig; NPU-Fragmentierung
Codec plus NPU als Silizium oder IP	SoC-Hersteller	Der AV2-Zyklus beginnt	Starke etablierte Anbieter
Evaluierung, Konformität und deterministische Inferenz	Wer neuronale Werkzeuge ausliefert	Bitgenauigkeit ist eine harte Anforderung	Kleiner Markt, außer im Paket
Standardbeiträge zu H.267	Lizenzgeber, große Plattformen	Auswertung der Einreichungen im Januar 2027	Langwierig, teuer, ungewiss

Skepsis, bis eine Behauptung vier Dinge nennt

Behauptung	Was die Belege sagen
„Neuronale Codecs schlagen VVC bereits.“	Im Labor. DCVC-RT spart 21 % gegenüber dem VVC-Referenzencoder; sein Decoder kostet etwa 175 Tausend MACs pro Pixel.
„KI-Codecs werden H.26x und AV1 in wenigen Jahren ersetzen.“	Nicht bei der Wiedergabe: H.267 Version 1 Ende 2029, Chips um 2031 Schätzung. Der Codec, der jetzt kommt, AV2, ist klassisch.
„MLVC spart 88 % Bandbreite.“	Gegenüber H.264 bei 360p. Gegenüber Hardware-HEVC: 75,5 % bei 360p und 65,4 % bei 540p, auf Microsofts eigenem Testdatensatz.
„Neuronale Texturen sparen 85–96 % Grafikspeicher.“	In Demoszenen. Das ausgelieferte Spiel gewann etwa 30 %, nur bei ausgewählten Assets.
„JPEG AI ist ein ISO-Standard, also werden Kameras und Browser ihn übernehmen.“	Kein Produkt mit standardmäßiger Unterstützung gefunden. Das Web übernahm JPEG XL, ein klassisches Format.

Fragen Sie nach Referenz, Metrik, Auflösung und Decodieraufwand. Eine Behauptung, die nicht alle vier nennen kann, ist ein Forschungsergebnis, kein Produkt.

Worauf zu achten ist, und offene Fragen

WEGMARKEN

Januar 2027	H.267-Auswertung: wie viele Einreichungen auf neuronale Werkzeuge setzen, zu welchen Kosten pro Pixel
Anfang 2027	Erste AV2-Webstreams bei YouTube; AV2-Decoder in angekündigten SoCs
Zweite Hälfte 2027	Erstes Testmodell von H.267
2027	MLVC über Peer-to-Peer-Anrufe in Teams hinaus
2027–2028	Ein erstes Produkt, das JPEG AI standardmäßig nutzt
Ende 2029	H.267 Version 1 startet den Chipzyklus

OFFENE FRAGEN

01 Exaktheit im großen Maßstab: bitgenaue Ganzzahldecodierung auf NPUs aller Hersteller – mit welchem Kompressionsverlust?

02 Speicher vor MACs: der DRAM-Verkehr pro Pixel bei neuronaler Nachverarbeitung

03 Die Kosten pro Pixel, ab denen sich ein neuronales Werkzeug in H.267 nicht mehr lohnt

04 Welche geschlossenen Systeme neben Anrufen zuerst neuronale End-to-End-Codecs nutzen

05 Wie man Wiedergabetreue misst, wenn ein Decoder Details erzeugt

Das sind Felder, die wir gemeinsam erkunden können, im technischen Austausch und in technischer Zusammenarbeit.

i@tanshuai.com · tanshuai.com