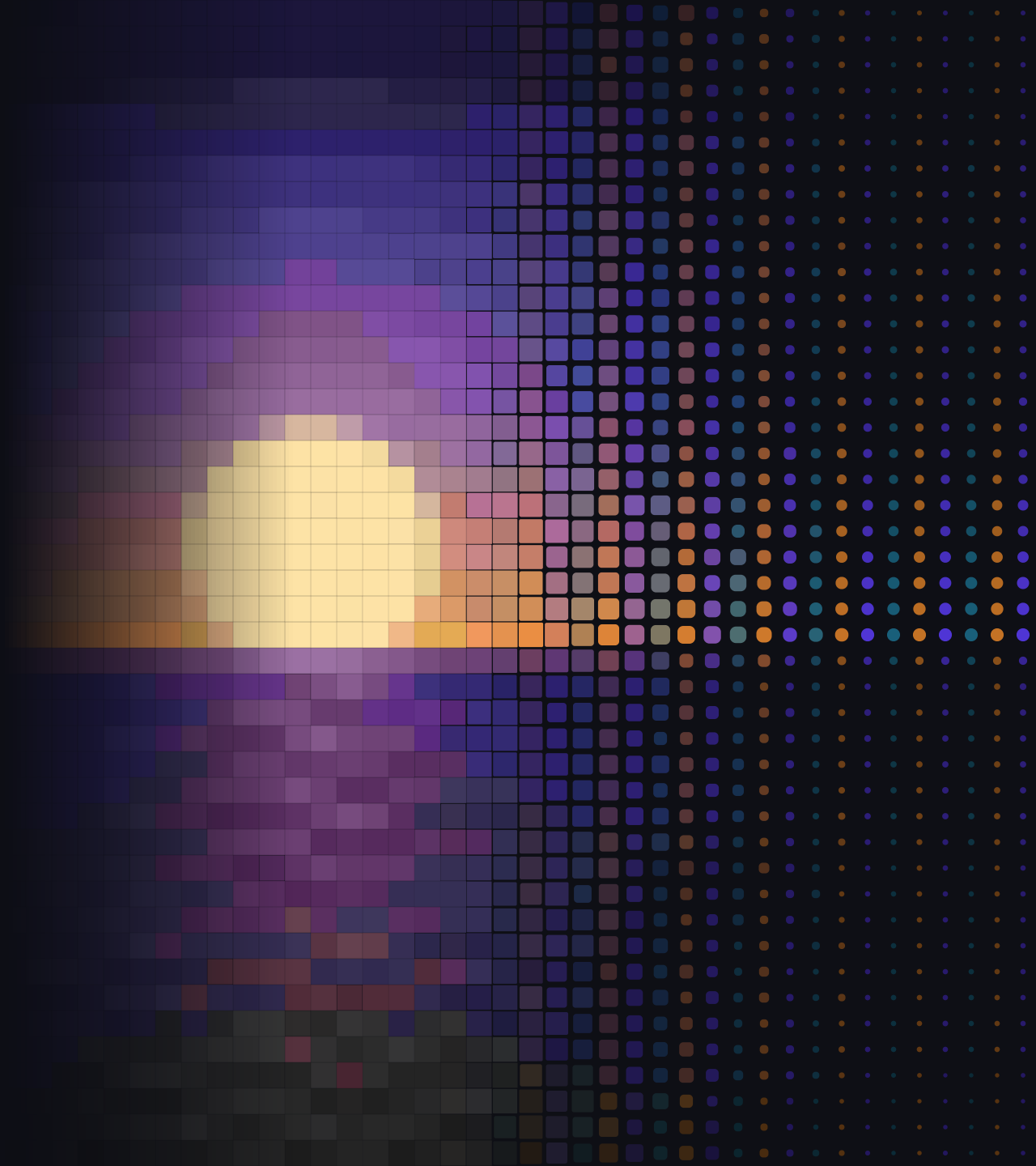
Forskningsnotat · november 2026

Nevral koding i 2026

Rundt kodeken, ikke i stedet for den

Avspilling · Foto · Samtaler · Generativ KI · Spill · Silisium

Tan Shuai
tanshuai.com · i@tanshuai.com



KONKLUSJONEN

Allerede utbredt, men rundt kodeken

01 Nevrale nettverk er utbredt i begge ender av standardkodekene

Kamera-ISP-er, nedskalering på kodingssiden og superoppløsning ved avspilling. Bitstrømmen i mellom er fortsatt AV1, HEVC, AVC eller VVC, dekodet av fastkoblet maskinvare.

02 Nevrale ende-til-ende-kodeker brukes bare i lukkede sløyfer

Samtaler, tokeniserere for generativ KI og ett AAA-spill: ett selskap kontrollerer begge ender og kan falle tilbake på en standardkodek.

03 Nevrale bitstrømmer på vanlige skjermer venter på H.267

Versjon 1 ventes ved utgangen av 2029; de første brikkene rundt 2031 og bred bruk midt på 2030-tallet **anslag**.

≈75 %

av mobil datatrafikk er video (utgangen av 2025)

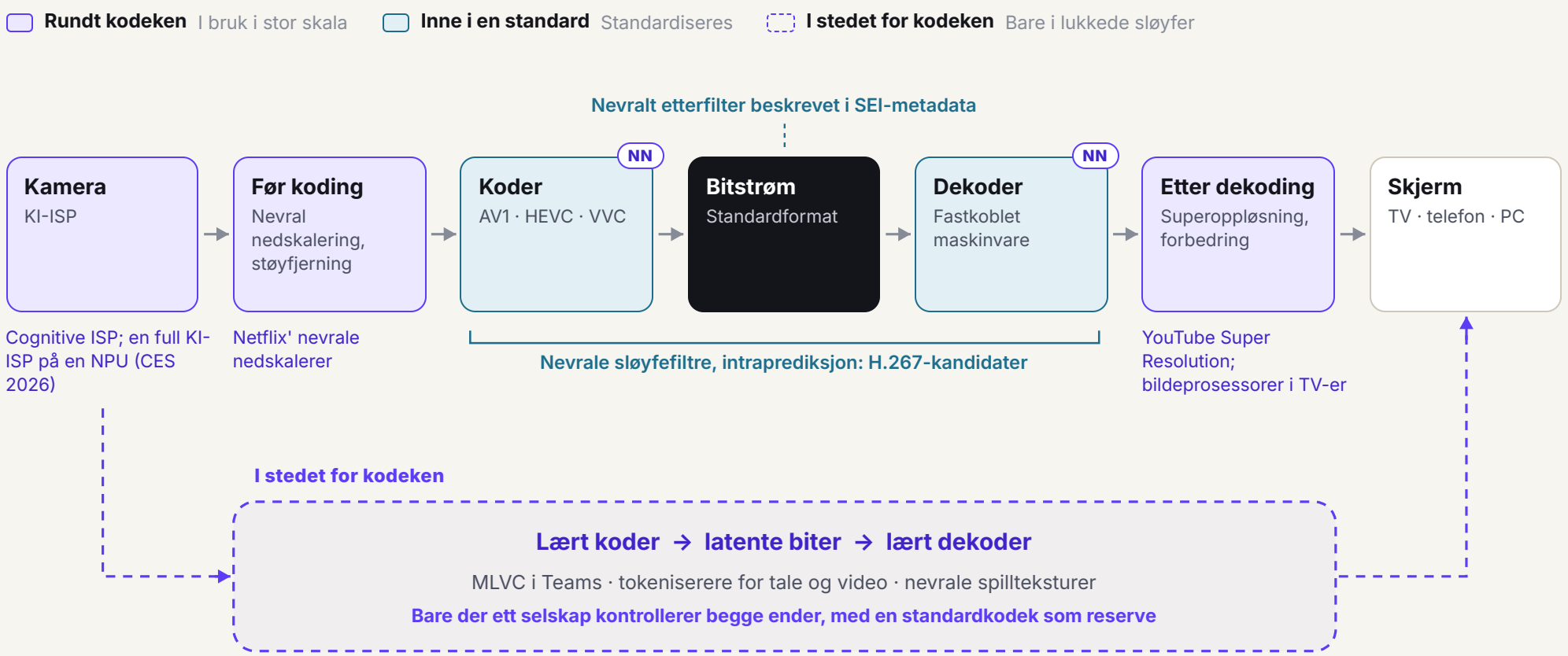
1,26 milliarder

smarttelefoner levert i 2025, og alle dekoder video i fastkoblet maskinvare

≈12×

så mye energi per piksel som en HEVC-dekoder i maskinvare, for det eneste publiserte designet av en nevral dekoderbrikke

Tre veier for nevrale nettverk inn i kodeken



> Rundt kodeken krever ingens tillatelse. Inne i en standard krever en standard. I stedet for kodeken krever kontroll over begge ender.

Hvor nevral koding tas i bruk i 2026

	Rundt en standardkodek forbehandling, forbedring, oppskalering	Nevral ende-til-ende-kodek lært koder og dekoder	Inne i en standard nevrale verktøy i standard bitstrøm eller API
01 **Avspilling og strømming**	**I STOR SKALA** Netflix' nedskalerer; superoppløsning hos YouTube og Meta; KI-oppskalering i TV-er	**IKKE I BRUK** Ingen strømmetjeneste sender et nevralt videoformat; TV-er og telefoner dekoder standardkodeker i maskinvare	**STANDARDISERES** H.267 versjon 1 ventes ved utgangen av 2029; nevrale verktøy er blant kandidatene
02 **Foto og bildebehandling**	**I STOR SKALA** KI-ISP-er i flaggskiptelefoner; en full nevral ISP vist som IP på CES 2026	**DEMOER, PILOTER** JPEG AI er en ISO-standard: telefondemoer, ingen produkter	**IKKE I BRUK** Formatet som vinner terreng, er klassisk: JPEG XL på som standard i Chrome 155
03 **Samtaler og møter**	**I STOR SKALA** Superoppløsning hos mottakeren og nevral støydemping	**I STOR SKALA** Nevral lyd i Teams-samtaler mellom to parter siden 2021; MLVC-video rulles ut i Teams	**I BRUK, BEGRENSET** Opus 1.5 legger til valgfrie, bakoverkompatible ML-verktøy
04 **Generativ KI**	—	**I STOR SKALA** Alle video- og talemodeller kjører en lært tokeniserer, f.eks. en video-VAE på 1:192	—
05 **Spill og grafikk**	**I STOR SKALA** Nevral oppskalering og bildegenerering: DLSS 4 alene i 250+ spill og apper	**I BRUK, BEGRENSET** Nevrale teksturer tatt i bruk i ett AAA-spill; NVIDIAs SDK er i beta	**STANDARDISERES** DirectX får matriseoperasjoner for shadere (forhåndsversjon)
06 **Spesialsamband: satellitt, roboter**	**DEMOER, PILOTER** Gjenoppretting av strømmer med lav bitrate på serversiden	**DEMOER, PILOTER** Tapsrobuste nevrale kodeker i forskning og piloter	—

Radene er ordnet etter hvor mange mennesker hvert område når.

Nevralt i begge ender, standard i midten

- **Ved koderen**

  Netflix' nevrale nedskalerer har kjørt siden 2022 uten endringer på noen enhet; omtrent 77 % av seerne foretrakk den i Netflix' tester leverandørpåstand.

- **Ved skjermen**

  YouTube Super Resolution på TV-er siden oktober 2025; Meta kjører superoppløsning på serverne og på telefoner.

- **I TV-en**

  Samsungs prosessor NQ8 AI Gen3 markedsføres med 768 nevrale nettverk for oppskalering.

BITSTRØMMEN FORBLIR STANDARD

30 %

av seertiden på Netflix i AV1 (desember 2025)

>90 %

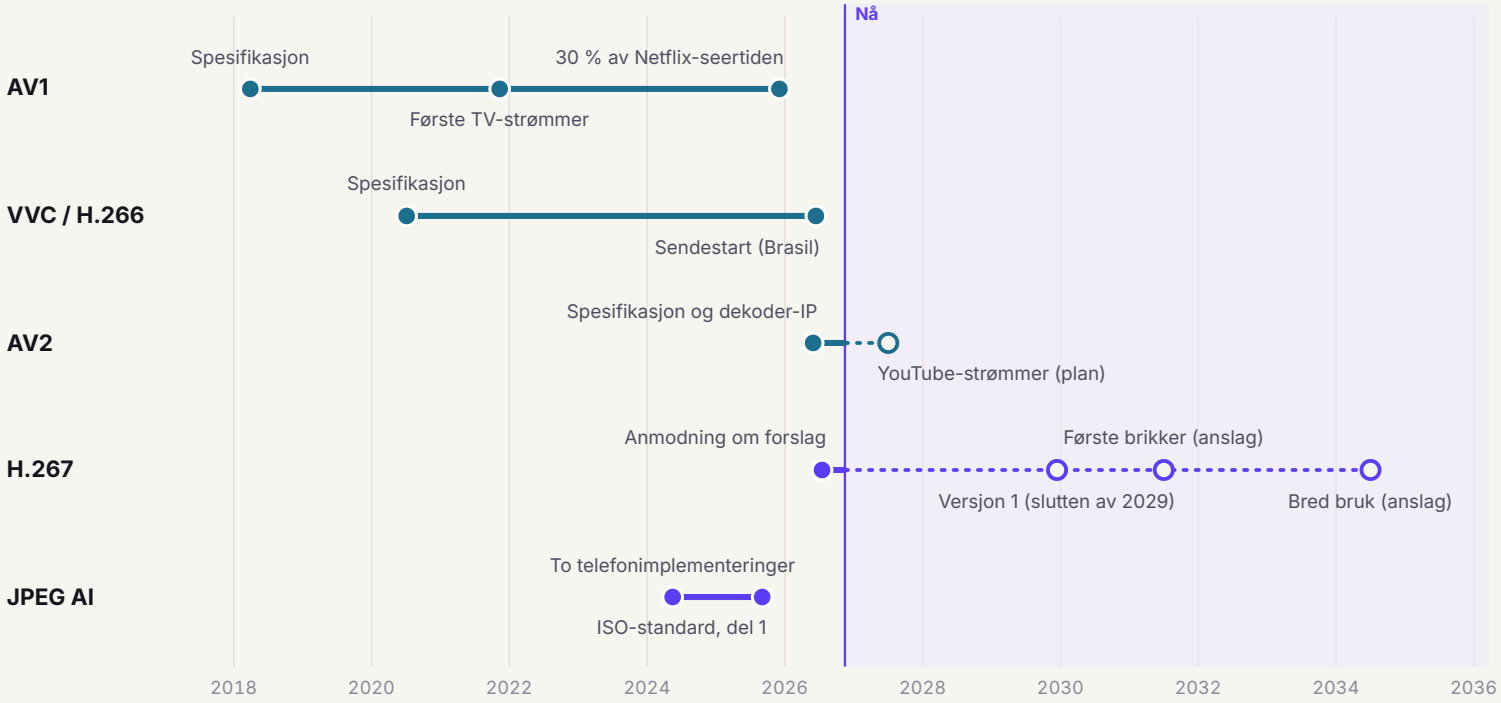
av YouTubes katalog, målt i seertid, tilgjengelig i AV1 (mars 2026)

3 uker

fra AV2 versjon 1.0.0 (28. mai 2026) til kunder for dekoder-IP hos tre leverandører

TIDSLINJE

Fra ferdig standard til vanlige skjermer går det store deler av et tiår



> AV1 brukte nesten åtte år fra spesifikasjonen (2018) til 30 % av seertiden på Netflix (desember 2025). Åpne markører er mål eller mine anslag.

H.267 avgjør om nevrale verktøy når alle skjermer

Oktober 2025
Anmodning om dokumentasjon: fem svar fra 16 selskaper og laboratorier; nevrale gevinster tydeligst på 8K HDR

Juli 2026
Anmodning om forslag sendt ut i Genève

Januar 2027
Svarene vurderes i Brisbane

Andre halvår 2027
Første testmodell

Slutten av 2029
Versjon 1 av standarden

≈28 %
spart mot VVC av det klassiske ECM-sporet

6–14 %
spart av nevrale sløyfefiltre ved 10 til 22+ tusen MAC per piksel

SEI
VVC-strømmer kan allerede signalisere et nevralt etterfilter som dekoderne kan bruke

Min vurdering: en klassisk kjerne med noen få nevrale verktøy der bitene som spares, forsvarer dekoderkostnaden analyse **.**

Nevralt inne i kameraet, klassisk på nettet

- **Kamera**

  Qualcomms Cognitive ISP har kjørt segmentering i sanntid i telefoner siden 2023; en full KI-ISP som kjører på en NPU, ble vist på CES 2026.

- **JPEG AI**

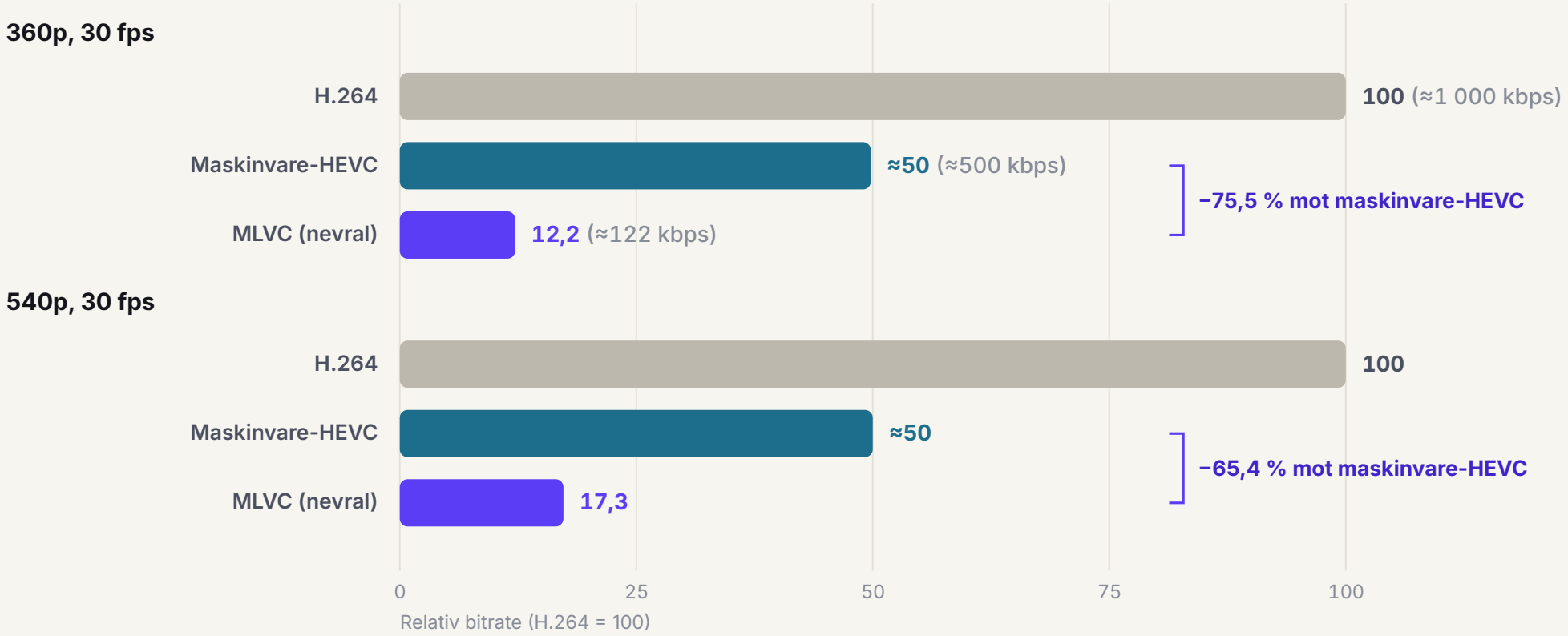
  Publisert i 2025: 12,5–27,9 % bedre enn VVC intra ved 7 til 216 tusen MAC per piksel. Jeg fant ingen produkter som bruker den som standard.

- **Nettet**

  Chrome 155 slo på JPEG XL som standard 6. oktober 2026, med en minnesikker dekoder skrevet i Rust.

> **Formater som må kunne åpnes overalt, forblir klassiske; nevral koding vinner inne i ett selskaps egen kjede.**

Samtaler er der nevrale ende-til-ende-kodeker brukes



Satin har kjørt i alle samtaler mellom to parter i Teams og Skype siden 2021. MLVC, åpen kildekode under MIT siden 24. juli 2026, rulles ut i peer-to-peer-samtaler i Teams i opptil 540p30 og bruker høyst halvparten av NPU-en **leverandørpåstand**.

Ikke alle plattformer valgte nevralt, og grunnene er lærerike

INNE I EN STANDARD, VALGFRITT

Opus 1.5 og 1.6

Nevral forbedring, maskering av pakketap, dyp redundans og båndbreddeutvidelse, alt avslått som standard og fullt kompatibelt med standarden fra 2012.

KLASSISK, MED VILJE

Meta MLow (2024)

Mer enn 20 % av Metas samtaler går på ARMv7-telefoner, og titalls millioner WhatsApp-samtaler per dag på telefoner som er over ti år gamle. MLow gir omtrent dobbelt så god kvalitet som Opus ved 6 kbps med 10 % mindre regnekraft.

Kompleksitetsbudsjettet i enhetsparken avgjør. Nevrale kodeker tas først i bruk der enhetene er nye og én leverandør kontrollerer begge ender.

Den største utrullingen som ingen kaller en kodek

- **Lærte kodeker overalt**

  Alle tale- og videomodeller leser og skriver tokens eller latente verdier fra en lært kodek.

- **Ikke distribusjonsformater**

  Hver modell har sitt eget latentrom: ingen ratekontroll, ingen konformitetstest, ingen annen dekoder.

- **En kostnadspost**

  For alle som leverer generativ video, kutter en mer komprimerende latent representasjon genereringskostnaden mer enn proporsjonalt analyse.

TO EKSEMPLER

12,5 Hz · 1,1 kbps

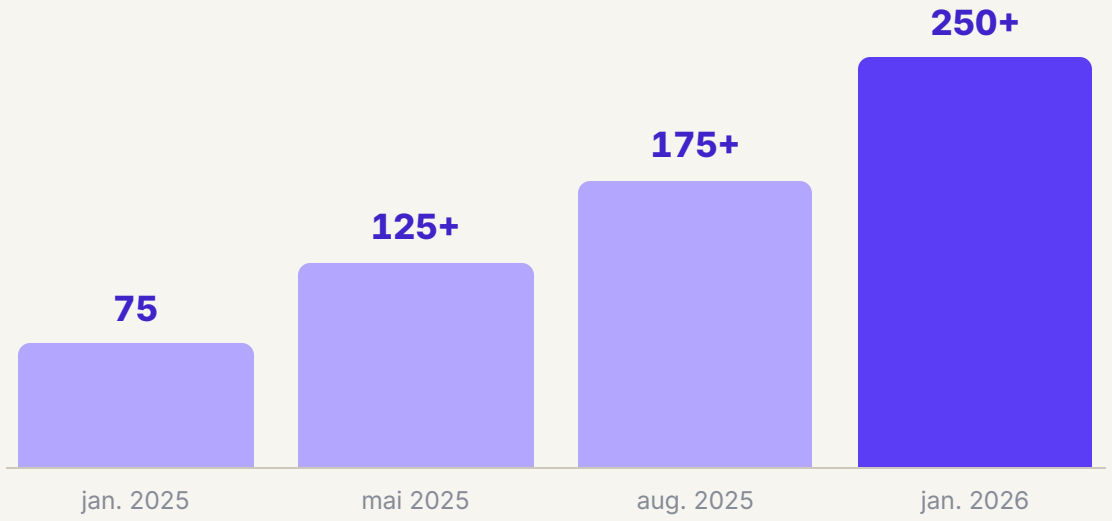
taletokens fra Kyutais Mimi, lavere enn noen telefonikodek

1:192

video-VAE-en i LTX-Video: ett latent token for 32×32 piksler over 8 bilder

Nevral rekonstruksjon er utbredt; nevrale ressurser er på vei inn

DLSS 4 Multi Frame Generation: spill og apper



Nevral teksturkomprimering

NVIDIAs demoscene, GTC 2026

Teksturminne 6,5 GB → 0,97 GB

85 % mindre minne

Assassin's Creed Mirage, utgitt

Bare utvalgte ressurser; mot blokkformater

≈30 % bedre komprimering

Ulike mål: ikke direkte sammenlignbare

DirectX Linear Algebra kom i offentlig forhåndsversjon i april 2026: én kodebane for nevral dekoding i shadere på alle leverandørers GPU-er.

SPESIALSAMBAND

Satellitter, droner og roboter: best egnet, et lite marked

- **Hvorfor det passer**

  Én eier kontrollerer begge ender, båndbredden er knapp, og innholdet er avgrenset nok til å trene for.

- **Kina**

  Kunngjøring nr. 49 av 2026, gjeldende fra 18. november, reviderer eksportkontrollen for droner; én oppføring bruker semantisk bildekomprimering som terskel.

- **USA**

  FCC førte 22. desember 2025 utenlandsproduserte droner og kritiske komponenter til dem opp på sin Covered List.

En solid nisje for en spesialist som kan håndtere eksportregelverket, ikke et massemarked.

En størrelsesorden mer energi per piksel

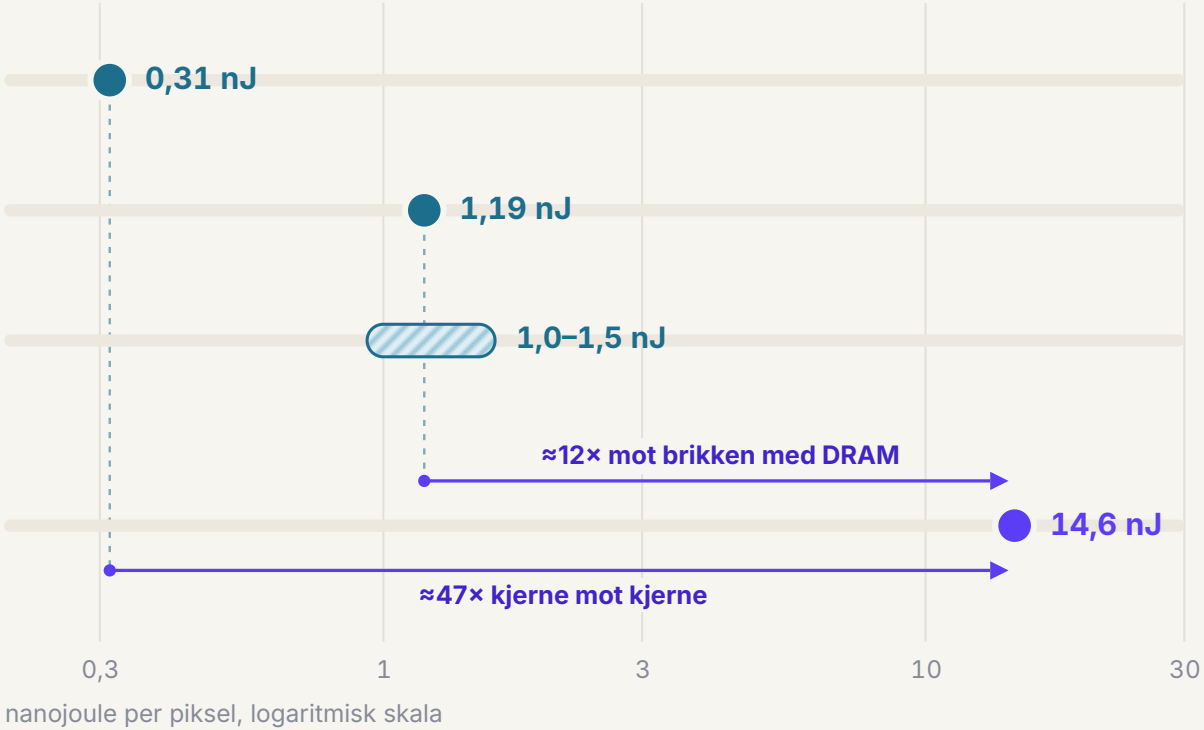
HEVC-dekoderbrikke, kjerne
40 nm, 4K30, målt

HEVC-dekoderbrikke, med DRAM
40 nm, 4K30, målt kjerne pluss simulert DRAM

VVC-dekoder
28 nm, 4K60, fremskrivning

Nevral videodekoder (NVCA)
28 nm, 1080p25, kun syntese; vekter og DRAM ikke medregnet

Ved 1080p30 er det omtrent 74 mW mot 908 mW.



Ved 1080p30: omtrent 74 mW for HEVC-brikken med DRAM mot rundt 900 mW for NVCA. Eldre node for HEVC-brikken, vekter og DRAM ikke medregnet for NVCA: sammenligningen favoriserer det nevrale designet.

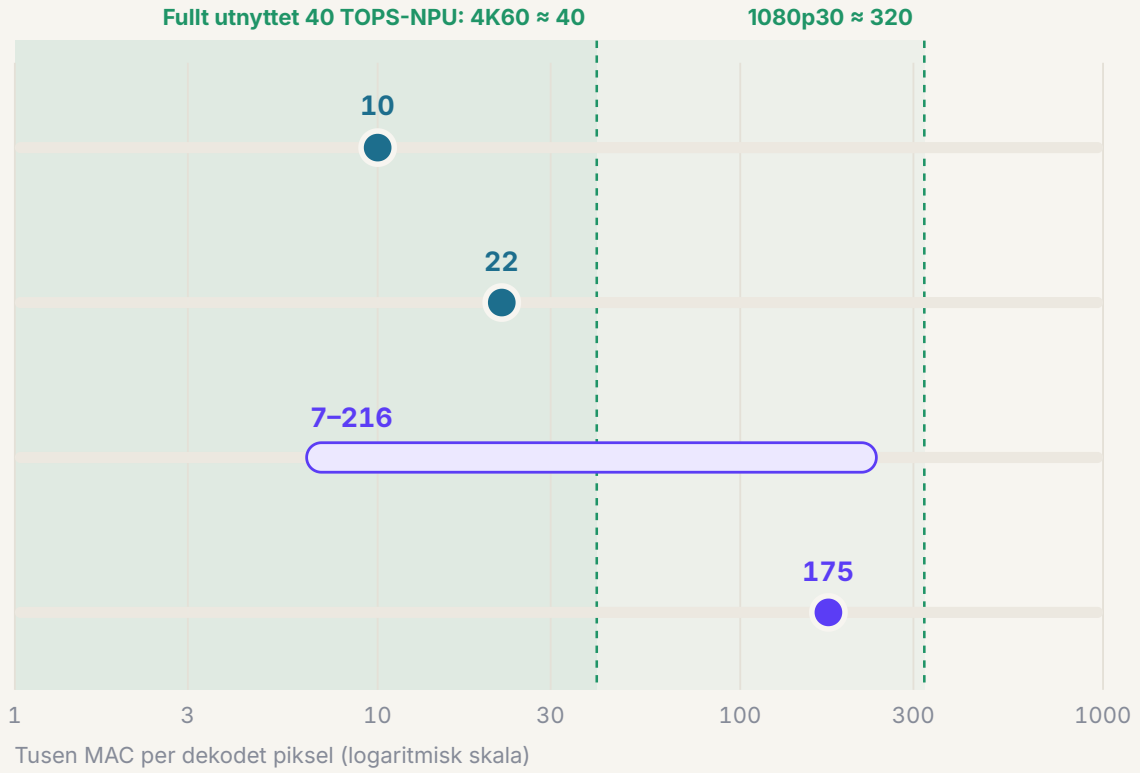
Hva en nevral dekoder krever av en NPU

Nevralt sløyfefilter, svært lav kompleksitet
JVET-utforskning · ≈6 % bitbesparelse
1,2 TOPS ved 1080p30 · 10 ved 4K60

Nevralt sløyfefilter, lav kompleksitet
JVET-utforskning · ≈8 %
2,7 TOPS ved 1080p30 · 22 ved 4K60

JPEG AI-dekodere, basis til høy
12,5–27,9 % mot VVC intra

DCVC-RT, en hel nevral videodekoder
21 % mot VVCs referanse, artikkelens oppsett
22 TOPS ved 1080p30 · 174 ved 4K60



> DCVC-RT trenger omtrent 175 tusen MAC per piksel: omtrent 22 TOPS ved 1080p30 og 174 ved 4K60, ved full utnyttelse **anslag**. En fullt utnyttet NPU på 40 TOPS har råd til omtrent 40 tusen ved 4K60.

Minnetrafikk og bitnøyaktighet er de harde grensene

DRAM DOMINERER

Minne før MAC

I MITs HEVC-brikke står DRAM for omtrent tre firedeler av systemets energi: 0,88 av 1,19 nJ per piksel. Nevrale dekodere bærer store egenskapskart mellom bildene, og forfatterne av DCVC-RT fant at minne-I/O, ikke aritmetikk, begrenset hastigheten.

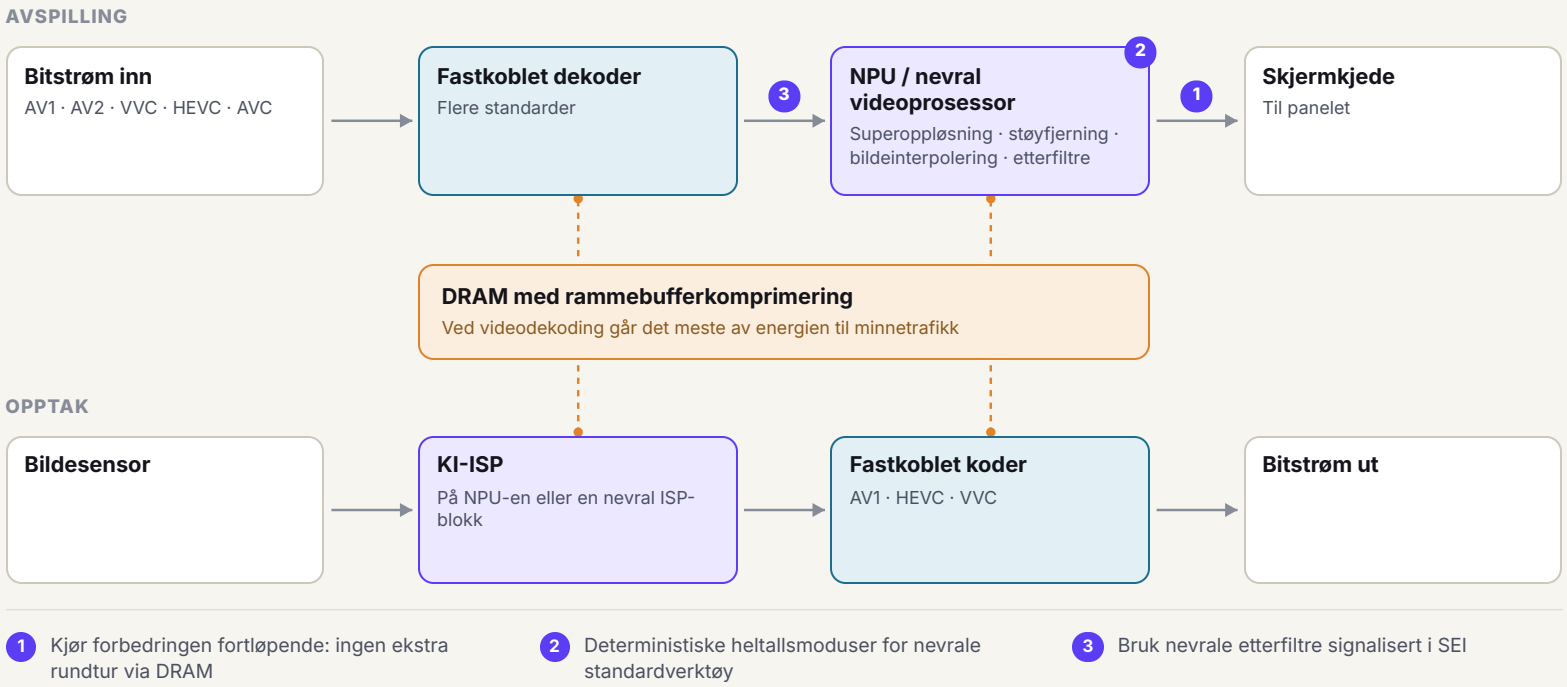
EKSAKTHET

Alle dekodere må beregne de samme bitene

Entropidekodingen feiler hvis to enheter beregner en sannsynlighet ulikt. Løsningene: heltallsmodeller (DCVC-RT), en hyperprior med delt skala (MLVC), 16-biters heltallsinferens (JVETs SADL), konformitetspunkter (JPEG AI).

For et brikketeam er det HEVCs disiplin med heltallstransformer på nytt: standarden må definere aritmetikken nøyaktig.

Arbeidsdelingen i dagens produkter: en fastkoblet kodek pluss en NPU



> Tencents Canghai: mer enn 100 000 brikker av første generasjon i drift; andre generasjon legger til H.266 og superoppløsning på brikken. Chips&Media og Allegro DVT selger kodek-IP og nevral video-IP side om side.

Hva jeg ville bygget, etter rolle

SoC- og TV-brikketeam

- Fastkoblede dekodere for alle standarder du må spille av, AV2 nå
- En NPU dimensjonert for forbedring i skjermoppløsning, uten ekstra rundtur via DRAM
- SEI for nevrale etterfiltre og deterministiske heltallsmoduser
- H.267 som designstart i 2029 eller senere

IP-leverandører

- AV2- og VVC-dekodere nå
- Akseleratorer for etterfiltre og rammebufferkomprimering
- Konformitetsverktøy for nevral prosessering

Videoplattformer

- Nevral forbehandling ved koderen lønner seg først
- Superoppløsning på klientsiden der enhetene tillater det
- Nevrale ende-til-ende-kodeker bare i lukkede sløyfer du kontrollerer

Hvor nevrale verktøy blir lisensierbare

Standard	Nevralt innhold	Status, november 2026
H.267 (arbeidsnavn)	Åpent: klassiske (ECM) og nevrale verktøy er begge kandidater	Svarene vurderes i januar 2027; versjon 1 ved utgangen av 2029
VVC og HEVC, med VSEI	Nevralt etterfilter beskrevet i SEI; valgfritt for dekodere	I kraft; utvidet i oktober 2025
AV2	Klassisk, royaltyfri	Versjon 1.0.0 den 28. mai 2026; YouTube-strømmer på nettet planlagt tidlig i 2027
JPEG AI	Lært ende-til-ende-bildekodek	Del 1 publisert 2025; fant ingen produkter med støtte som standard
IEEE 1857.11	Nevral bildekoding designet for NPU-er	Publisert desember 2024
Video Coding for Machines	Koding for maskinanalyse i stedet for visning	Utkast til internasjonal standard tidlig i 2026
ML-utvidelser i Opus	Valgfrie og bakoverkompatible	I libopus 1.5 og 1.6

645

patentfamilier for VVCs etterfølger eid av ByteDance (GreyB, 2025)

0,50 · 1,00–1,50 dollar

VVC-poolens sats per telefon og per 4K-TV, Region 1

0 dollar

i royalty for AV1 og AV2 fra AOMedias medlemmer

HVEM FÅR BETALT

Pengene følger standardene og plattformene

TJENER NÅ

Plattformer

Beholder besparelsene

Netflix' nevrale nedskalerer; superoppløsning hos YouTube og Meta; Microsoft ga ut MLVC som åpen kildekode under MIT

TJENER NÅ

Enhets- og brikkeprodusenter

Selger differensiering

Kamera-ISP-er; oppskaleringsprosessorer i TV-er; Tencents egne Canghai-kodekbrikker

TJENER NÅ

IP-leverandører

Lisensierer hver nye standard

AV2-dekoderkunder hos Chips&Media, Allegro DVT og VeriSilicon innen tre uker etter spesifikasjonen

TJENER NÅ

GPU-leverandører

Bygger plattforminnlåsing

DLSS 4 i 250+ spill og apper; SDK for nevrale teksturer i beta

PER ENHET

Patenthavere og -pooler

Tar betalt per enhet

VVC: 0,50 dollar per telefon, 1,00–1,50 dollar per 4K-TV; H.267-verktøy etter 2029

EXIT VED OPPKJØP

Oppstartsselskaper innen nevrale kodeker

Kjøpt opp

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

> De tre mest kjente oppstartsselskapene innen nevrale kodeker, WaveOne, iSIZE og Deep Render, ble kjøpt av Apple, Sony og InterDigital.

HVOR ET SELSKAP HAR MULIGHETER

Seks muligheter, og risikoen i hver

Mulighet	Hvem kjøper	Hvorfor nå	Hovedrisiko
Nevral forbehandling ved koderen	Strømme- og brukervideoplattformer, kringkastere	Ingen endring på enhetene; bevist hos Netflix	Store plattformer bygger det selv
Forbedrings-IP og finjustering på skjermsiden	Produsenter av TV-, TV-boks- og telefonbrikker	NPU-ene finnes allerede i enhetene	Kommodifisering; interne team
Nevrale kodeker i lukkede sløyfer	Leverandører av bedriftskommunikasjon, industri og telepresence	MLVC viser at det virker, og koden er åpen	Du må kontrollere begge ender; NPU-fragmentering
Silisium eller IP med kodek pluss NPU	SoC-produsenter	AV2-syklusen er i gang	Sterke etablerte aktører
Evaluering, konformitet og deterministisk inferens	Alle som leverer nevrale verktøy	Bitnøyaktighet er et absolutt krav	Lite marked uten pakkeløsning
Standardbidrag til H.267	Lisensgivere, store plattformer	Svarene vurderes i januar 2027	Langt, kostbart, usikkert

Ta påstander med en klype salt til de oppgir fire ting

Påstand	Hva dokumentasjonen viser
«Nevrale kodeker slår allerede VVC.»	I laboratoriet. DCVC-RT sparer 21 % mot VVCs referansekoder; dekoderen koster omtrent 175 tusen MAC per piksel.
«KI-kodeker vil erstatte H.26x og AV1 innen få år.»	Ikke i avspilling: H.267 versjon 1 ved utgangen av 2029, brikker rundt 2031 **anslag**. Kodeken som kommer nå, AV2, er klassisk.
«MLVC kutter båndbredden med 88 %.»	Mot H.264 ved 360p. Mot maskinvare-HEVC: 75,5 % ved 360p og 65,4 % ved 540p, på Microsofts eget testsett.
«Nevrale teksturer kutter videominnet med 85–96 %.»	I demoscener. Det utgitte spillet fikk omtrent 30 %, bare på utvalgte ressurser.
«JPEG AI er en ISO-standard, så kameraer og nettlesere vil ta den i bruk.»	Fant ingen produkter som støtter den som standard. Nettet tok i bruk JPEG XL, et klassisk format.

Spør etter referansen, kvalitetsmålet, oppløsningen og dekodingskostnaden. En påstand som ikke kan oppgi alle fire, er et forskningsresultat, ikke et produkt.

VERDT Å FØLGE

Hva du bør følge med på, og åpne spørsmål

MILEPÆLER

Januar 2027	H.267-svarene vurderes: hvor mange bygger på nevrale verktøy, og til hvilken kostnad per piksel
Tidlig i 2027	YouTubes første AV2-strømmer på nettet; AV2-dekodere i annonserte SoC-er
Andre halvår 2027	H.267s første testmodell
2027	MLVC utover peer-to-peer-samtaler i Teams
2027–2028	Et første produkt som bruker JPEG AI som standard
Slutten av 2029	H.267 versjon 1 starter brikkesyklusen

ÅPNE SPØRSMÅL

01 Eksakthet i stor skala: bitnøyaktig heltallsdekoding på ulike leverandørers NPU-er, med hvilket komprimeringstap?

02 Minne før MAC: DRAM-trafikken per piksel for nevral etterbehandling

03 Kostnaden per piksel der et nevralt verktøy slutter å lønne seg i H.267

04 Hvilke lukkede sløyfer som følger etter samtaler med nevrale ende-til-ende-kodeker

05 Hvordan måle troskap mot originalen når en dekoder genererer detaljer

Dette er områder vi kan utforske sammen, gjennom teknisk utveksling og teknisk samarbeid.

i@tanshuai.com · tanshuai.com