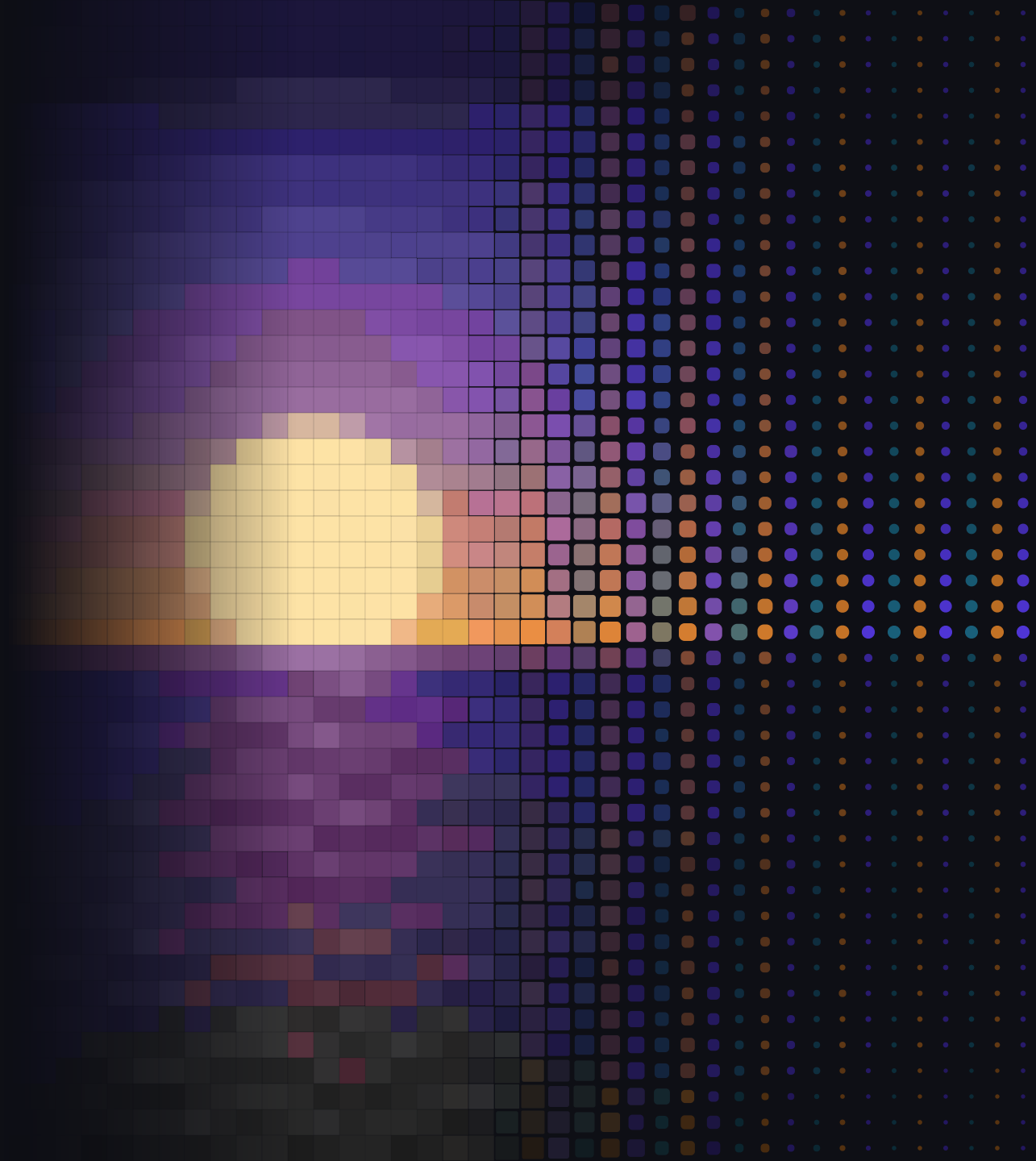
Research note · November 2026

Neural coding in 2026

Around the codec, not instead of it

Playback · Photography · Calls · Generative AI · Games · Silicon

Tan Shuai
tanshuai.com · i@tanshuai.com



THE VERDICT

Mainstream already, but around the codec

01 Neural networks are mainstream at both ends of standard codecs

Camera ISPs, encode-side downscaling and playback super-resolution. The bitstream in between is still AV1, HEVC, AVC or VVC, decoded by fixed-function hardware.

02 End-to-end neural codecs ship only in closed loops

Calls, generative-AI tokenizers and one AAA game: one company controls both ends and can fall back to a standard codec.

03 Neural bitstreams on mainstream screens wait for H.267

Version 1 is due at the end of 2029; first chips around 2031 and mainstream use in the mid-2030s estimate.

≈75%

of mobile data traffic is video (end of 2025)

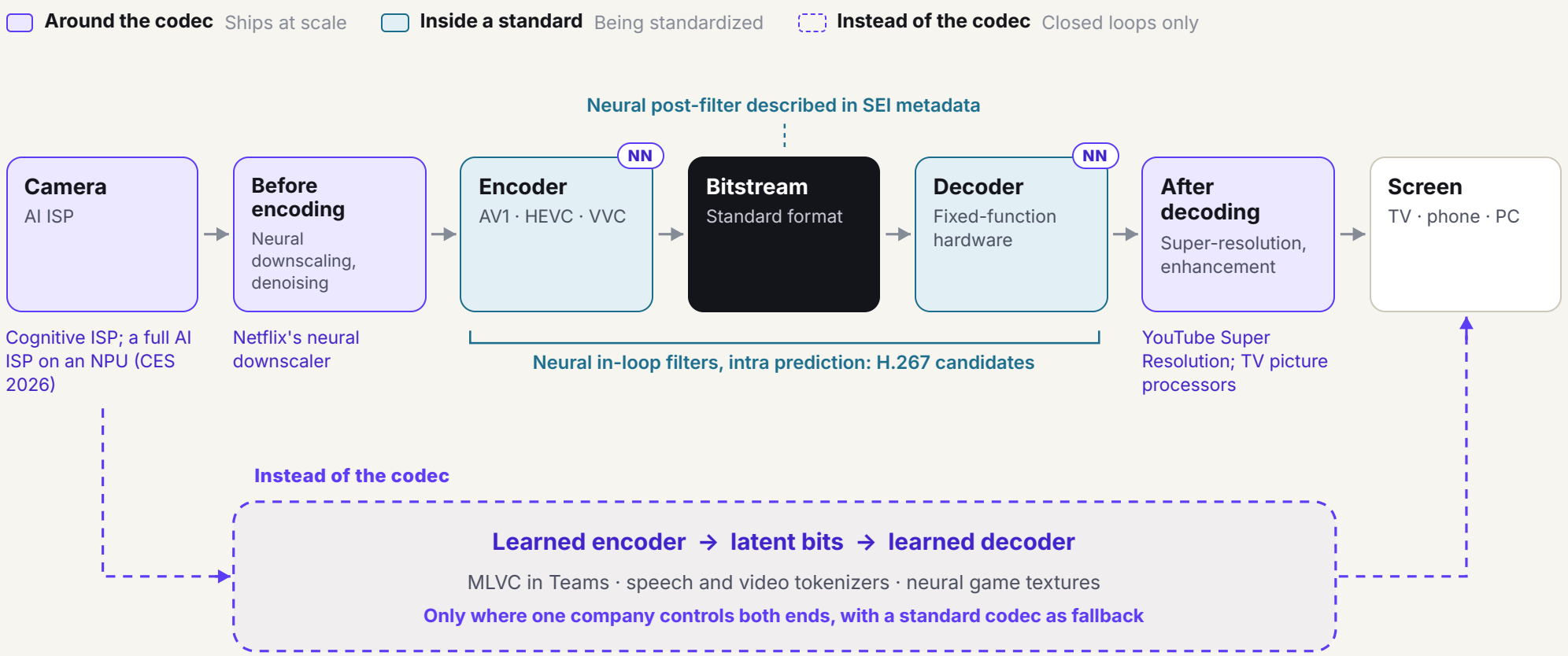
1.26 billion

smartphones shipped in 2025, all decoding video in fixed-function hardware

≈12×

the energy per pixel of a hardware HEVC decoder, for the only published neural decoder chip design

Three ways a neural network enters a codec



> Around the codec needs no one's permission. Inside a standard needs a standard. Instead of the codec needs control of both ends.

Where neural coding ships in 2026

	Around a standard codec pre-processing, enhancement, upscaling	End-to-end neural codec learned encoder and decoder	Inside a standard neural tools in a standard bitstream or API
01 **Playback and streaming**	AT SCALE Netflix downscaler; YouTube and Meta super-resolution; TV AI upscalers	NOT DEPLOYED No streaming service sends a neural video format; TVs and phones decode standard codecs in hardware	STANDARDIZING H.267 version 1 is due at the end of 2029; neural tools are among the candidates
02 **Photography and imaging**	AT SCALE AI ISPs in flagship phones; a full neural ISP shown as IP at CES 2026	DEMOS, PILOTS JPEG AI is an ISO standard: phone demos, no product	NOT DEPLOYED The format gaining ground is classical: JPEG XL on by default in Chrome 155
03 **Calls and meetings**	AT SCALE Receiver-side super-resolution and neural noise suppression	AT SCALE Neural audio in Teams two-party calls since 2021; MLVC video rolling out in Teams	SHIPPING, LIMITED Opus 1.5 adds optional, backward-compatible ML tools
04 **Generative AI**	—	AT SCALE Every video and voice model runs a learned tokenizer, e.g. a 1:192 video VAE	—
05 **Games and graphics**	AT SCALE Neural upscaling and frame generation: DLSS 4 alone in 250+ games and apps	SHIPPING, LIMITED Neural textures shipped in one AAA game; NVIDIA's SDK is in beta	STANDARDIZING DirectX is adding matrix operations for shaders (preview)
06 **Specialist links: satellite, robots**	DEMOS, PILOTS Server-side restoration of low-bitrate feeds	DEMOS, PILOTS Loss-resilient neural codecs in research and pilots	—

Rows are ordered by how many people each domain reaches.

Neural at both ends, standard in the middle

- **At the encoder**

  Netflix's neural downscaler has run since 2022 with no change on any device; about 77% of viewers preferred it in Netflix's tests vendor claim.

- **At the screen**

  YouTube Super Resolution on TVs since October 2025; Meta runs super-resolution on its servers and on phones.

- **In the TV**

  Samsung's NQ8 AI Gen3 processor advertises 768 neural networks for upscaling.

THE BITSTREAM STAYS STANDARD

30%

of Netflix viewing in AV1 (December 2025)

90%+

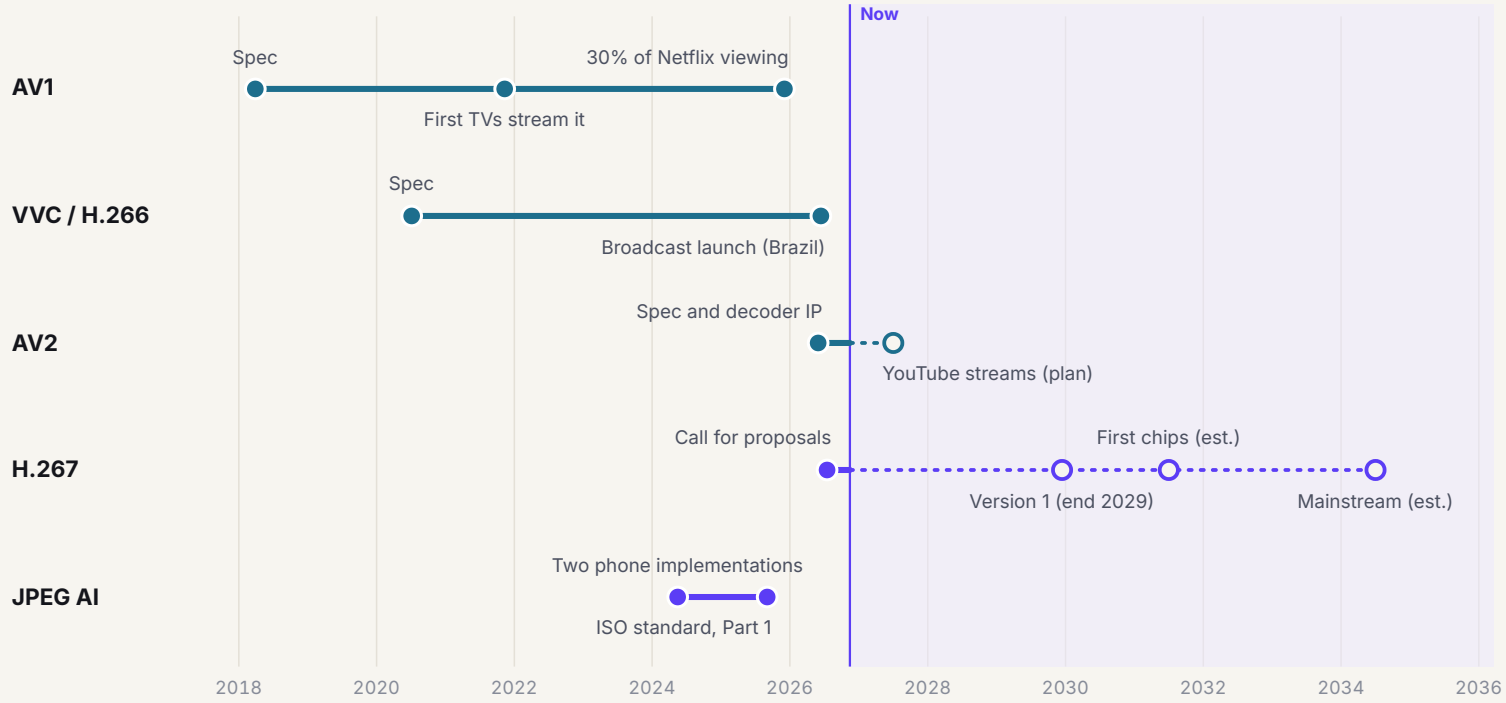
of YouTube's catalogue, by watch time, available in AV1 (March 2026)

3 weeks

from AV2 version 1.0.0 (May 28, 2026) to decoder IP customers at three vendors

TIMING

From a finished standard to mainstream screens takes most of a decade



AV1 took nearly eight years from its specification (2018) to 30% of Netflix viewing (December 2025). Hollow marks are targets or my estimates.

H.267 decides whether neural tools reach every screen

October 2025
Call for evidence: five responses from 16 companies and labs; neural gains clearest on 8K HDR

July 2026
Call for proposals issued in Geneva

January 2027
Responses evaluated in Brisbane

Second half of 2027
First test model

End of 2029
Version 1 of the standard

≈28%
saved over VVC by the classical ECM track

6–14%
saved by neural in-loop filters at 10 to 22+ thousand MACs per pixel

SEI
VVC streams can already signal a neural post-filter for decoders to apply

My reading: a classical core with a few neural tools where the bits saved justify the decoder cost analysis **.**

Neural inside the camera, classical on the web

- **Camera**

  Qualcomm's Cognitive ISP has run real-time segmentation in phones since 2023; a full AI ISP that runs on an NPU was shown at CES 2026.

- **JPEG AI**

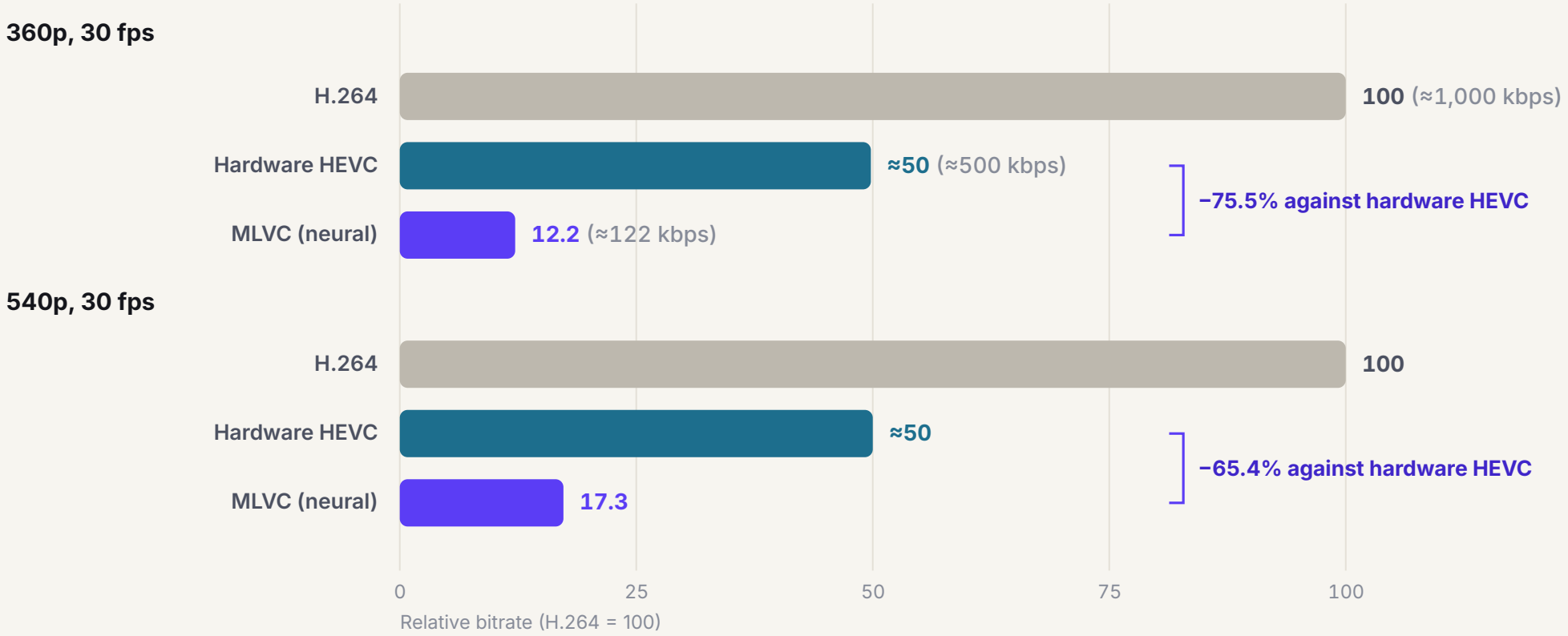
  Published in 2025: 12.5–27.9% over VVC intra at 7 to 216 thousand MACs per pixel. I found no product that uses it by default.

- **Web**

  Chrome 155 turned on JPEG XL by default on October 6, 2026, with a memory-safe decoder written in Rust.

Formats that must open everywhere stay classical; neural coding wins inside one company's pipeline.

Calls are where end-to-end neural codecs ship



Satin has run in all two-party Teams and Skype calls since 2021. MLVC, open source under MIT since July 24, 2026, is rolling out in peer-to-peer Teams calls at up to 540p30, using at most half of the NPU **vendor claim**.

CALLS AND MEETINGS

Not every platform went neural, and the reasons are instructive

INSIDE A STANDARD, OPTIONAL

Opus 1.5 and 1.6

Neural enhancement, packet-loss concealment, deep redundancy and bandwidth extension, all off by default and fully compatible with the 2012 standard.

CLASSICAL, BY CHOICE

Meta MLow (2024)

More than 20% of Meta's calls run on ARMv7 phones, and tens of millions of WhatsApp calls a day on phones over ten years old. MLow gives about twice Opus's quality at 6 kbps with 10% less compute.

The complexity budget of the installed base decides. Neural codecs ship first where devices are new and one vendor controls both ends.

GENERATIVE AI

The largest deployment nobody calls a codec

- **Learned codecs everywhere**

  Every voice and video model reads and writes the tokens or latents of a learned codec.

- **Not distribution formats**

  Each model has its own latent space: no rate control, no conformance test, no second decoder.

- **A cost line**

  For anyone serving generative video, a more compressive latent cuts generation cost more than proportionally analysis.

TWO EXAMPLES

12.5 Hz · 1.1 kbps

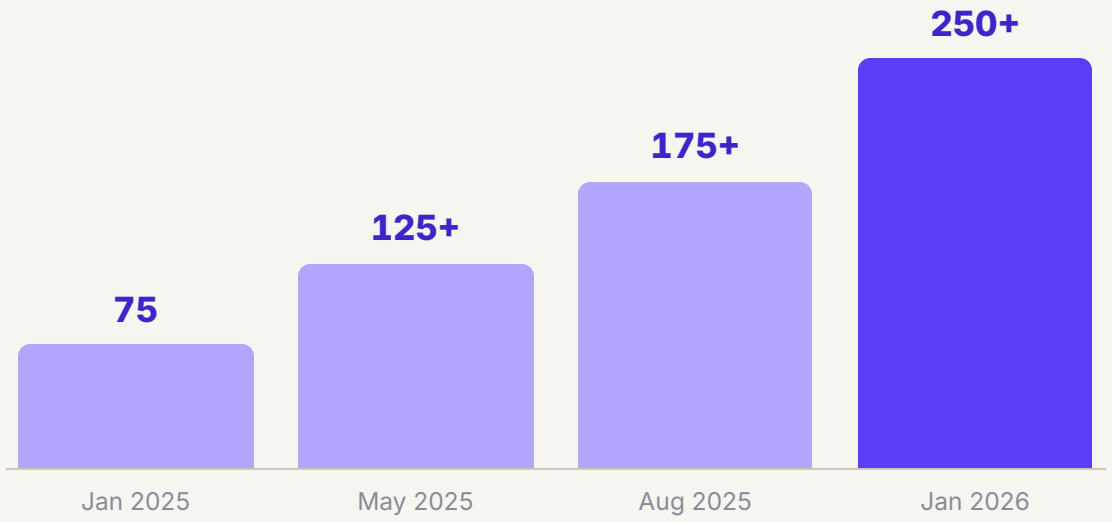
Kyutai's Mimi speech tokens, below any telephony codec

1:192

LTX-Video's video VAE: one latent token for 32×32 pixels over 8 frames

Neural reconstruction is mainstream; neural assets are starting

DLSS 4 Multi Frame Generation: games and apps



Neural texture compression

NVIDIA demo scene, GTC 2026

Texture memory 6.5 GB → 0.97 GB

85% less memory

Assassin's Creed Mirage, shipped

Selected assets only; against block formats

≈30% better compression

Different measures: not directly comparable

DirectX Linear Algebra entered public preview in April 2026: one code path for neural decoding inside shaders on any vendor's GPU.

Satellites, drones and robots: the best fit, a small market

- **Why it fits**

  One owner controls both ends, bandwidth is scarce, and the content is narrow enough to train for.

- **China**

  Announcement No. 49 of 2026, effective November 18, revises drone export controls; one entry uses semantic image compression as a threshold.

- **United States**

  The FCC added foreign-made drones and their critical components to its Covered List on December 22, 2025.

> **A sound niche for a specialist with export-compliance capacity, not a mainstream market.**

An order of magnitude more energy per pixel

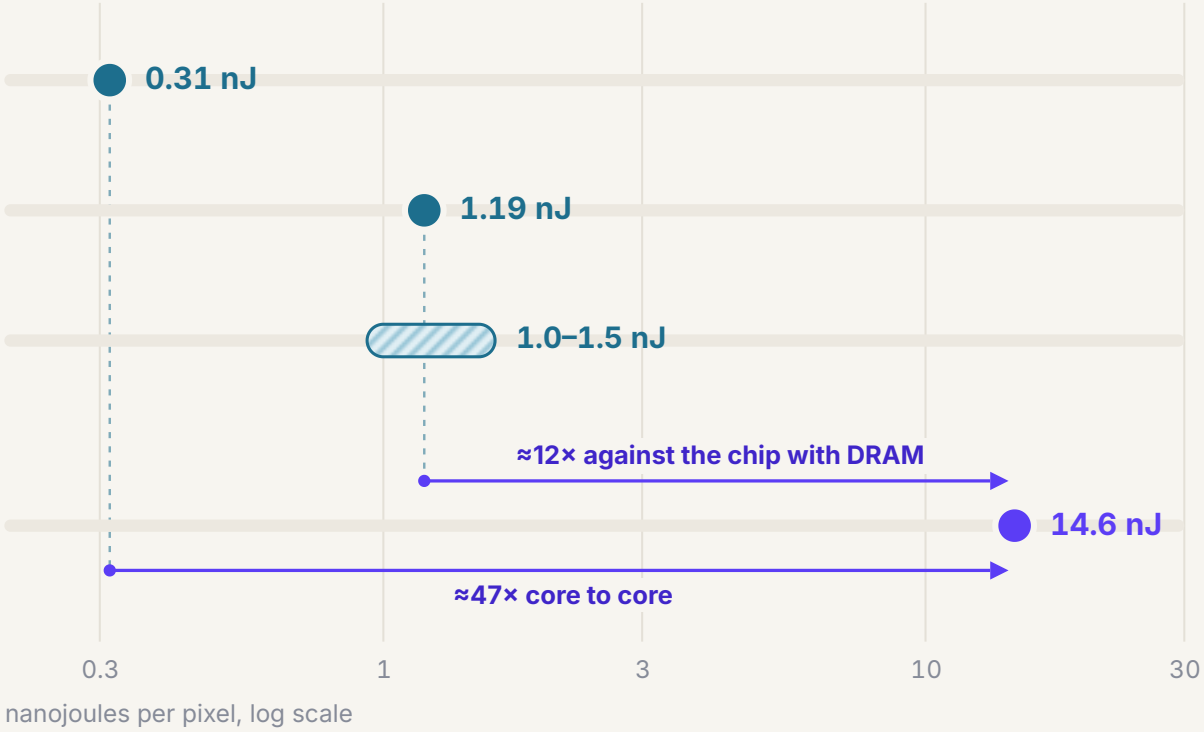
HEVC decoder chip, core
40 nm, 4K30, measured

HEVC decoder chip, with DRAM
40 nm, 4K30, measured core plus simulated DRAM

VVC decoder
28 nm, 4K60, projection

Neural video decoder (NVCA)
28 nm, 1080p25, synthesis only; weights and DRAM not counted

At 1080p30 that is about 74 mW against 908 mW.



At 1080p30: about 74 mW for the HEVC chip with DRAM against roughly 900 mW for NVCA. An older node for the HEVC chip, weights and DRAM not counted for NVCA: the comparison flatters the neural design.

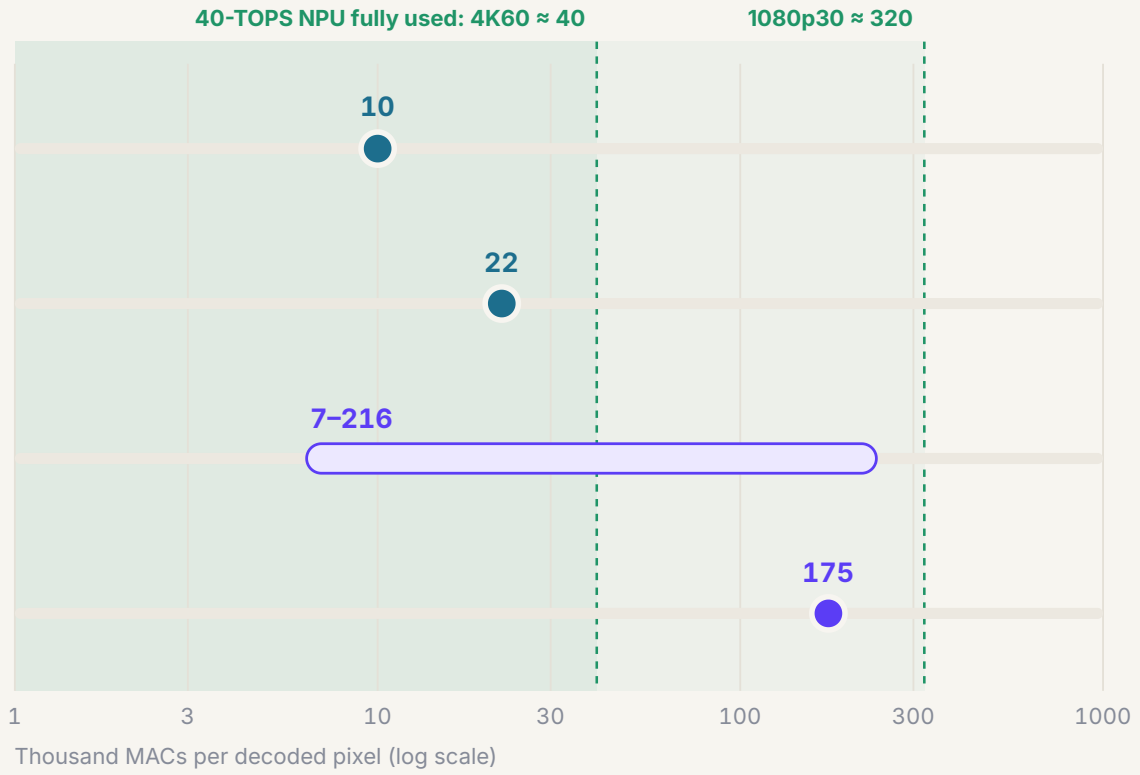
What a neural decoder asks of an NPU

Neural in-loop filter, very low complexity
JVET exploration · ≈6% bit saving
1.2 TOPS at 1080p30 · 10 at 4K60

Neural in-loop filter, low complexity
JVET exploration · ≈8%
2.7 TOPS at 1080p30 · 22 at 4K60

JPEG AI decoders, base to high
12.5–27.9% against VVC intra

DCVC-RT, a whole neural video decoder
21% against VVC's reference, paper setting
22 TOPS at 1080p30 · 174 at 4K60



DCVC-RT needs about 175 thousand MACs per pixel: about 22 TOPS at 1080p30 and 174 at 4K60, at full utilization estimate. A fully used 40-TOPS NPU affords about 40 thousand at 4K60.

Memory traffic and bit-exactness are the hard limits

DRAM DOMINATES

Memory before MACs

In the MIT HEVC chip, DRAM is about three-quarters of system energy: 0.88 of 1.19 nJ per pixel. Neural decoders carry large feature maps between frames, and DCVC-RT's authors found memory I/O, not arithmetic, limited speed.

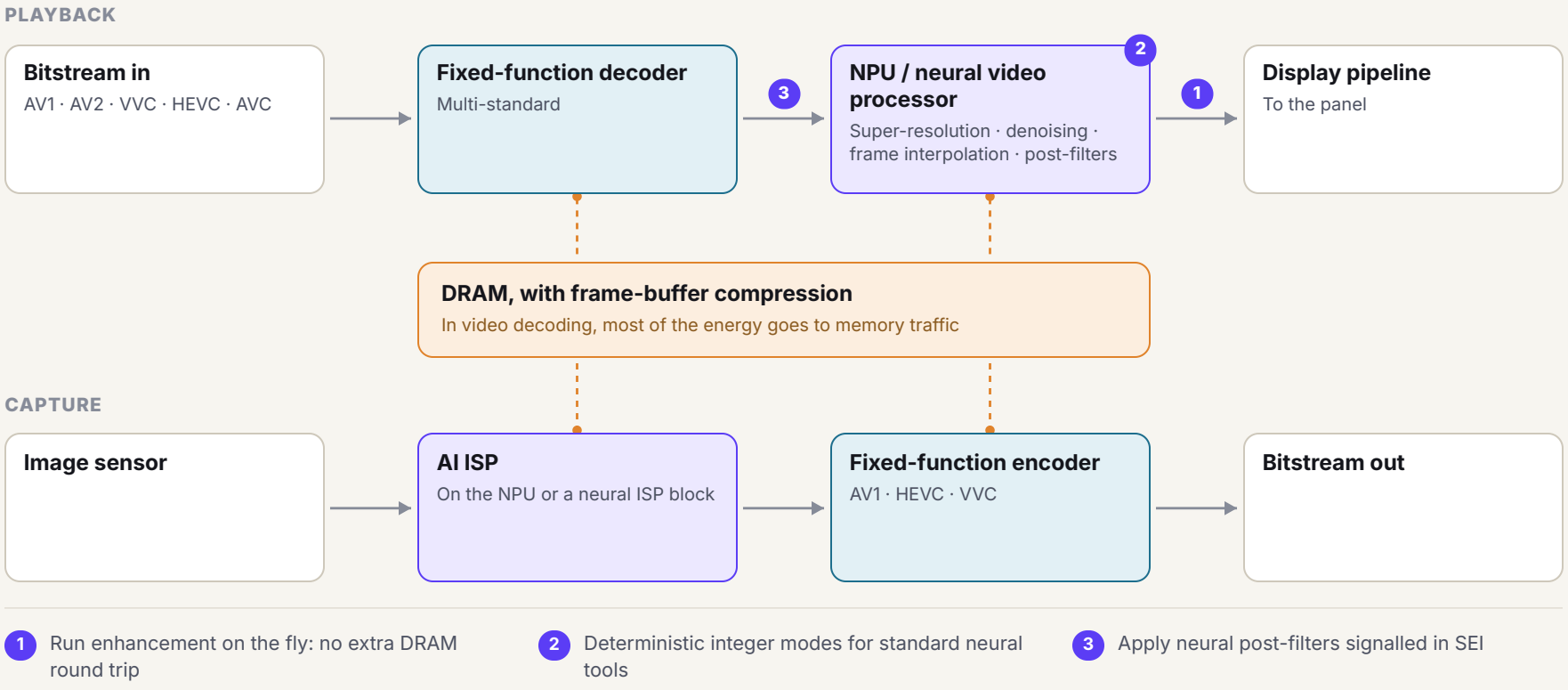
EXACTNESS

Every decoder must compute the same bits

Entropy decoding fails if two devices compute a probability differently. The fixes: integer models (DCVC-RT), a shared-scale hyperprior (MLVC), 16-bit integer inference (JVET's SADL), conformance points (JPEG AI).

For a chip team it is HEVC's integer-transform discipline again: the standard must define the arithmetic exactly.

The split that ships: a fixed-function codec plus an NPU



> Tencent's Canghai: more than 100,000 first-generation chips deployed; the second generation adds H.266 and on-chip super-resolution. Chips&Media and Allegro DVT sell codec IP and neural video IP side by side.

What I would build, by role

SoC and TV-chip teams

- Fixed-function decoders for every standard you must play, AV2 now
- An NPU sized for enhancement at display resolution, with no extra DRAM round trip
- Neural post-filter SEI and deterministic integer modes
- H.267 as a design start in 2029 or later

IP vendors

- AV2 and VVC decoders now
- Post-filter accelerators and frame-buffer compression
- Conformance tools for neural processing

Video platforms

- Neural preprocessing at the encoder pays back first
- Client-side super-resolution where devices allow
- End-to-end neural codecs only in closed loops you control

Where neural tools become licensable

Standard	Neural content	Status, November 2026
H.267 (working name)	Open: classical (ECM) and neural tools are both candidates	Responses evaluated January 2027; version 1 at the end of 2029
VVC and HEVC, with VSEI	Neural post-filter described in SEI; optional for decoders	In force; extended in October 2025
AV2	Classical, royalty-free	Version 1.0.0 on May 28, 2026; YouTube web streams planned for early 2027
JPEG AI	End-to-end learned image codec	Part 1 published 2025; no default product support found
IEEE 1857.11	Neural image coding designed for NPUs	Published December 2024
Video Coding for Machines	Coding for machine analysis rather than viewing	Draft international standard early 2026
Opus ML extensions	Optional and backward compatible	In libopus 1.5 and 1.6

645

beyond-VVC patent families held by ByteDance (GreyB, 2025)

$0.50 · $1.00–1.50

VVC pool rate per phone and per 4K TV, Region 1

$0

royalties for AV1 and AV2 from AOMedia's members

WHO GETS PAID

The money follows the standards and the platforms

EARNING NOW

Platforms

Keep the savings

Netflix's neural downscaler; YouTube and Meta super-resolution; Microsoft open-sourced MLVC under MIT

EARNING NOW

Device and chip makers

Sell differentiation

Camera ISPs; TV upscaling processors; Tencent's in-house Canghai codec chips

EARNING NOW

IP vendors

License each new standard

AV2 decoder customers at Chips&Media, Allegro DVT and VeriSilicon within three weeks of the spec

EARNING NOW

GPU vendors

Build platform lock-in

DLSS 4 in 250+ games and apps; neural texture SDK in beta

PER DEVICE

Patent holders and pools

Collect per device

VVC: $0.50 per phone, $1.00–1.50 per 4K TV; H.267 tools after 2029

EXIT BY ACQUISITION

Neural-codec startups

Acquired

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

> The three best-known neural-codec startups, WaveOne, iSIZE and Deep Render, were bought by Apple, Sony and InterDigital.

WHERE A COMPANY CAN PLAY

Six plays, and the risk in each

Play	Who buys	Why now	Main risk
Neural preprocessing at the encoder	Streaming and user-video platforms, broadcasters	No device change; proven at Netflix	Large platforms build it themselves
Display-side enhancement IP and tuning	TV, set-top and phone chip makers	NPUs are already in the devices	Commoditization; in-house teams
Closed-loop neural codecs	Enterprise communications, industrial and telepresence vendors	MLVC proves it works, and its code is open	You must control both ends; NPU fragmentation
Codec plus NPU silicon or IP	SoC makers	The AV2 cycle is starting	Strong incumbents
Evaluation, conformance and deterministic inference	Anyone shipping neural tools	Bit-exactness is a hard requirement	Small market unless bundled
Standards contributions to H.267	Licensors, large platforms	Responses are evaluated in January 2027	Long, costly, uncertain

CLAIMS YOU WILL HEAR

Discount a claim until it states four things

Claim	What the evidence says
"Neural codecs already beat VVC."	In the lab. DCVC-RT saves 21% against VVC's reference encoder; its decoder costs about 175 thousand MACs per pixel.
"AI codecs will replace H.26x and AV1 within a few years."	Not in playback: H.267 version 1 at the end of 2029, chips around 2031 estimate. The codec arriving now, AV2, is classical.
"MLVC cuts bandwidth by 88%."	Against H.264 at 360p. Against hardware HEVC: 75.5% at 360p and 65.4% at 540p, on Microsoft's own test set.
"Neural textures cut video memory by 85–96%."	In demo scenes. The shipped game gained about 30%, on selected assets only.
"JPEG AI is an ISO standard, so cameras and browsers will adopt it."	No default support found in products. The web adopted JPEG XL, a classical format.

Ask for the anchor, the metric, the resolution and the decode cost. A claim that cannot state all four is a research result, not a product.

WHAT TO WATCH

What to watch, and open questions

SIGNPOSTS

January 2027	H.267 responses evaluated: how many rely on neural tools, at what cost per pixel
Early 2027	YouTube's first AV2 web streams; AV2 decoders in announced SoCs
Second half of 2027	H.267's first test model
2027	MLVC beyond peer-to-peer Teams calls
2027–2028	A first product that uses JPEG AI by default
End of 2029	H.267 version 1 starts the silicon cycle

OPEN QUESTIONS

01 Exactness at scale: bit-exact integer decoding across vendors' NPUs, at what loss of compression?

02 Memory before MACs: the DRAM traffic per pixel of neural post-processing

03 The cost per pixel at which a neural tool stops paying for itself in H.267

04 Which closed loops beyond calls adopt end-to-end neural codecs first

05 How to measure fidelity when a decoder generates detail

These are areas we can explore together, through technical exchange and technical collaboration.

i@tanshuai.com · tanshuai.com