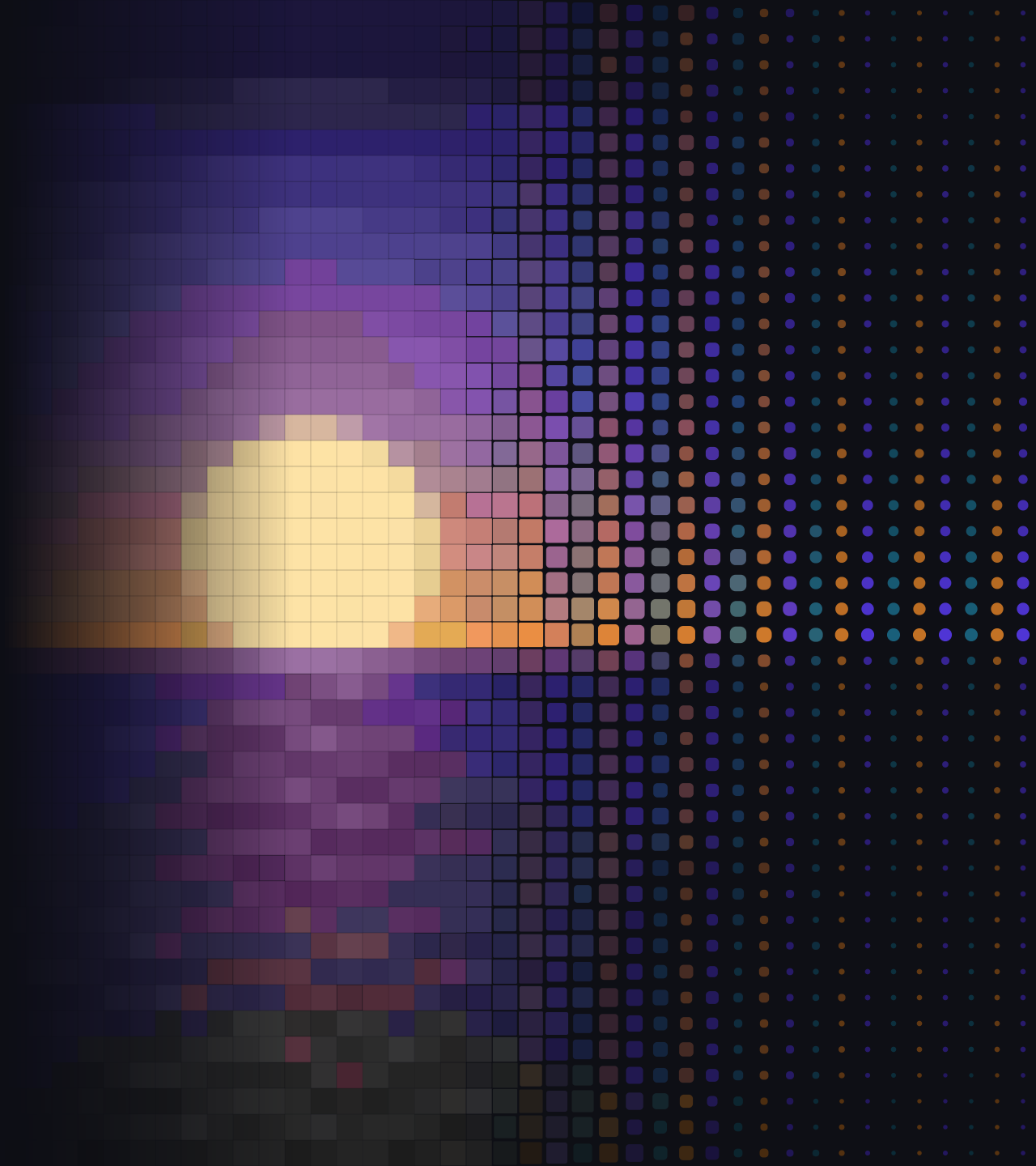
Nota de investigação · Novembro de 2026

Codificação neural em 2026

Em torno do codec, não em vez dele

Reprodução · Fotografia · Chamadas · IA generativa · Jogos · Silício

Tan Shuai
tanshuai.com · i@tanshuai.com



Já generalizada, mas em torno do codec

01 Redes neuronais: já generalizadas nas duas pontas dos codecs normalizados

ISP das câmaras, redução de resolução na codificação e super-resolução na reprodução. Pelo meio, o fluxo de bits continua a ser AV1, HEVC, AVC ou VVC, descodificado por hardware de função fixa.

02 Codecs neurais de ponta a ponta: só em circuitos fechados

Chamadas, tokenizadores de IA generativa e um jogo AAA: uma empresa controla as duas pontas e pode recorrer a um codec normalizado.

03 Fluxos de bits neurais nos ecrãs do grande público: à espera do H.267

Versão 1 prevista para o final de 2029; primeiros chips por volta de 2031 e uso generalizado em meados da década de 2030 estimativa.

≈75%

do tráfego de dados móveis é vídeo (final de 2025)

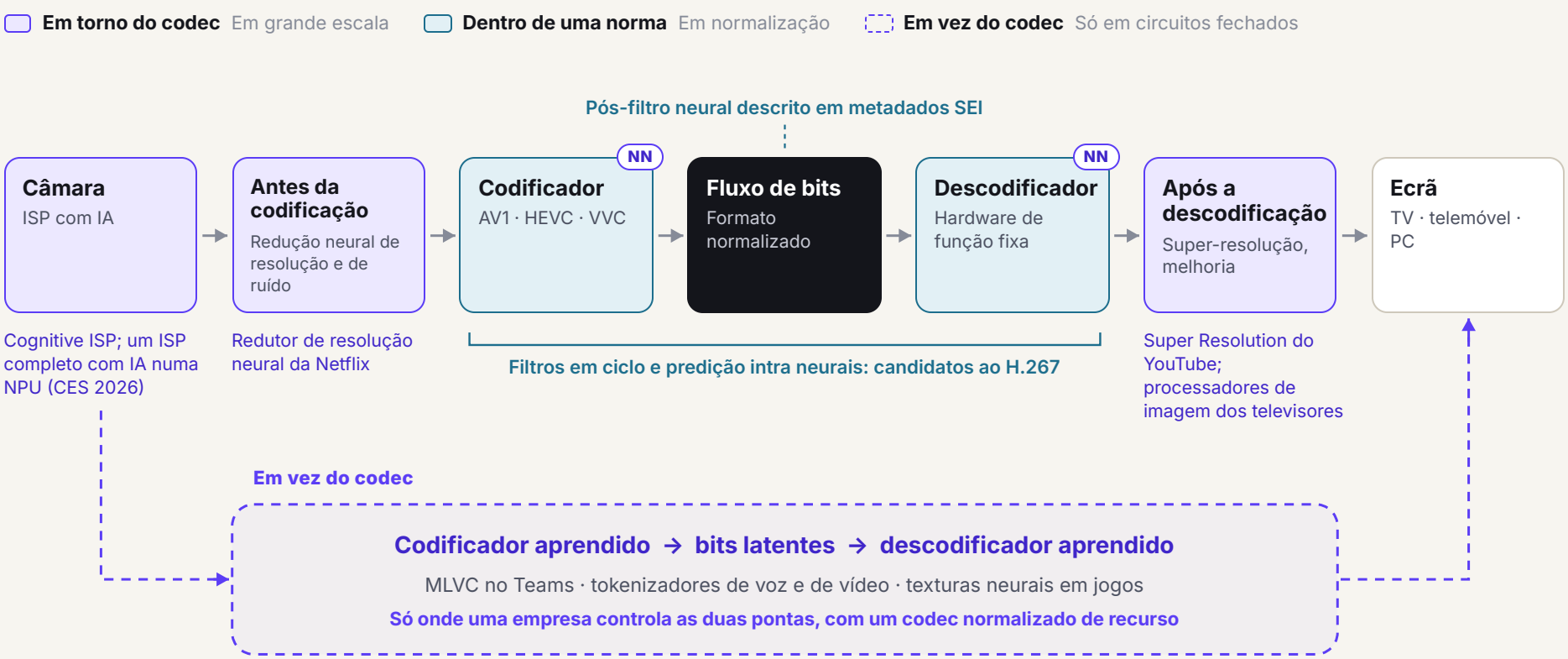
1,26 mil milhões

de smartphones entregues em 2025, todos a descodificar vídeo em hardware de função fixa

≈12×

a energia por píxel de um descodificador HEVC em hardware, no único projeto publicado de chip descodificador neural

Três formas de uma rede neuronal entrar num codec



> Em torno do codec, não é preciso autorização de ninguém. Dentro de uma norma, é preciso uma norma. Em vez do codec, é preciso controlar as duas pontas.

Onde a codificação neural chega ao mercado em 2026

	Em torno de um codec normalizado pré-processamento, melhoria, ampliação	Codec neural de ponta a ponta codificador e descodificador aprendidos	Dentro de uma norma ferramentas neurais num fluxo de bits ou API normalizados
01 **Reprodução e streaming**	**EM GRANDE ESCALA** Redutor de resolução da Netflix; super-resolução do YouTube e da Meta; ampliação por IA nos televisores	**NÃO IMPLANTADO** Nenhum serviço de streaming envia um formato de vídeo neural; televisores e telemóveis descodificam codecs normalizados em hardware	**EM NORMALIZAÇÃO** A versão 1 do H.267 está prevista para o final de 2029; há ferramentas neurais entre as candidatas
02 **Fotografia e imagem**	**EM GRANDE ESCALA** ISP com IA em telemóveis topo de gama; um ISP neural completo mostrado como IP na CES 2026	**DEMONSTRAÇÕES, PILOTOS** O JPEG AI é uma norma ISO: demonstrações em telemóveis, nenhum produto	**NÃO IMPLANTADO** O formato que ganha terreno é clássico: JPEG XL ativo por predefinição no Chrome 155
03 **Chamadas e reuniões**	**EM GRANDE ESCALA** Super-resolução no recetor e supressão neural de ruído	**EM GRANDE ESCALA** Áudio neural nas chamadas a dois do Teams desde 2021; vídeo MLVC em implementação no Teams	**USO LIMITADO** O Opus 1.5 acrescenta ferramentas de ML opcionais e retrocompatíveis
04 **IA generativa**	—	**EM GRANDE ESCALA** Cada modelo de vídeo e de voz usa um tokenizador aprendido, p. ex. um VAE de vídeo de 1:192	—
05 **Jogos e gráficos**	**EM GRANDE ESCALA** Ampliação neural e geração de fotogramas: só o DLSS 4 está em 250+ jogos e aplicações	**USO LIMITADO** Texturas neurais já lançadas num jogo AAA; o SDK da NVIDIA está em beta	**EM NORMALIZAÇÃO** O DirectX acrescenta operações matriciais aos shaders (pré-visualização)
06 **Ligações especializadas: satélite, robôs**	**DEMONSTRAÇÕES, PILOTOS** Restauro, no servidor, de fluxos de baixo débito	**DEMONSTRAÇÕES, PILOTOS** Codecs neurais resistentes a perdas, em investigação e pilotos	—

Linhas ordenadas pelo número de pessoas que cada domínio alcança.

Neural nas duas pontas, normalizado no meio

No codificador

A redução de resolução neural da Netflix funciona desde 2022 sem alterar nenhum dispositivo; cerca de 77% dos espetadores preferiram-na nos testes da Netflix **alegação do fornecedor**.

No ecrã

Super Resolution do YouTube nos televisores desde outubro de 2025; a Meta usa super-resolução nos servidores e nos telemóveis.

No televisor

O processador NQ8 AI Gen3 da Samsung anuncia 768 redes neuronais para ampliação.

O FLUXO DE BITS CONTINUA NORMALIZADO

30%

das visualizações da Netflix em AV1 (dez. 2025)

90%+

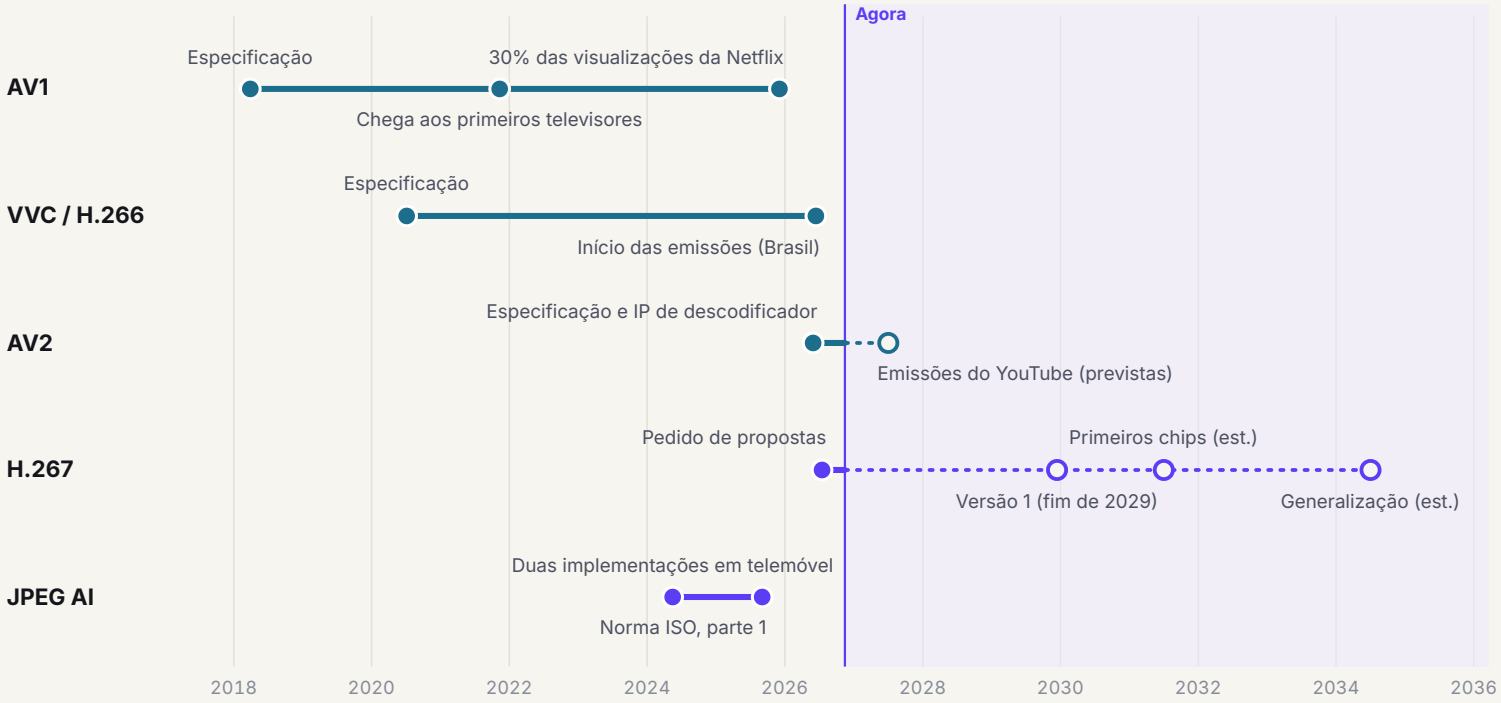
do catálogo do YouTube, por tempo de visualização, disponível em AV1 (mar. 2026)

3 semanas

do AV2 1.0.0 (28 de maio de 2026) a clientes de descodificadores em três fornecedores de IP

CALENDÁRIO

Entre uma norma concluída e os ecrãs do grande público passa a maior parte de uma década



O AV1 levou quase oito anos da especificação (2018) aos 30% das visualizações da Netflix (dezembro de 2025). Os marcadores vazios são metas ou estimativas minhas.

O H.267 decide se as ferramentas neurais chegam a todos os ecrãs

Outubro de 2025
Pedido de evidências: cinco respostas de 16 empresas e laboratórios; ganhos neurais mais claros em 8K HDR

Julho de 2026
Pedido de propostas lançado em Genebra

Janeiro de 2027
Respostas avaliadas em Brisbane

Segundo semestre de 2027
Primeiro modelo de teste

Final de 2029
Versão 1 da norma

≈28%
poupados face ao VVC pela via clássica, o ECM

6–14%
poupados por filtros em ciclo neurais, entre 10 e mais de 22 mil MAC por píxel

SEI
os fluxos VVC já podem sinalizar um pós-filtro neural para os descodificadores aplicarem

A minha leitura: um núcleo clássico com algumas ferramentas neurais onde os bits poupados justifiquem o custo no descodificador análise.

Neural dentro da câmara, clássico na Web

- **Câmara**

  O Cognitive ISP da Qualcomm faz segmentação em tempo real nos telemóveis desde 2023; na CES 2026 foi mostrado um ISP completo com IA que corre numa NPU.

- **JPEG AI**

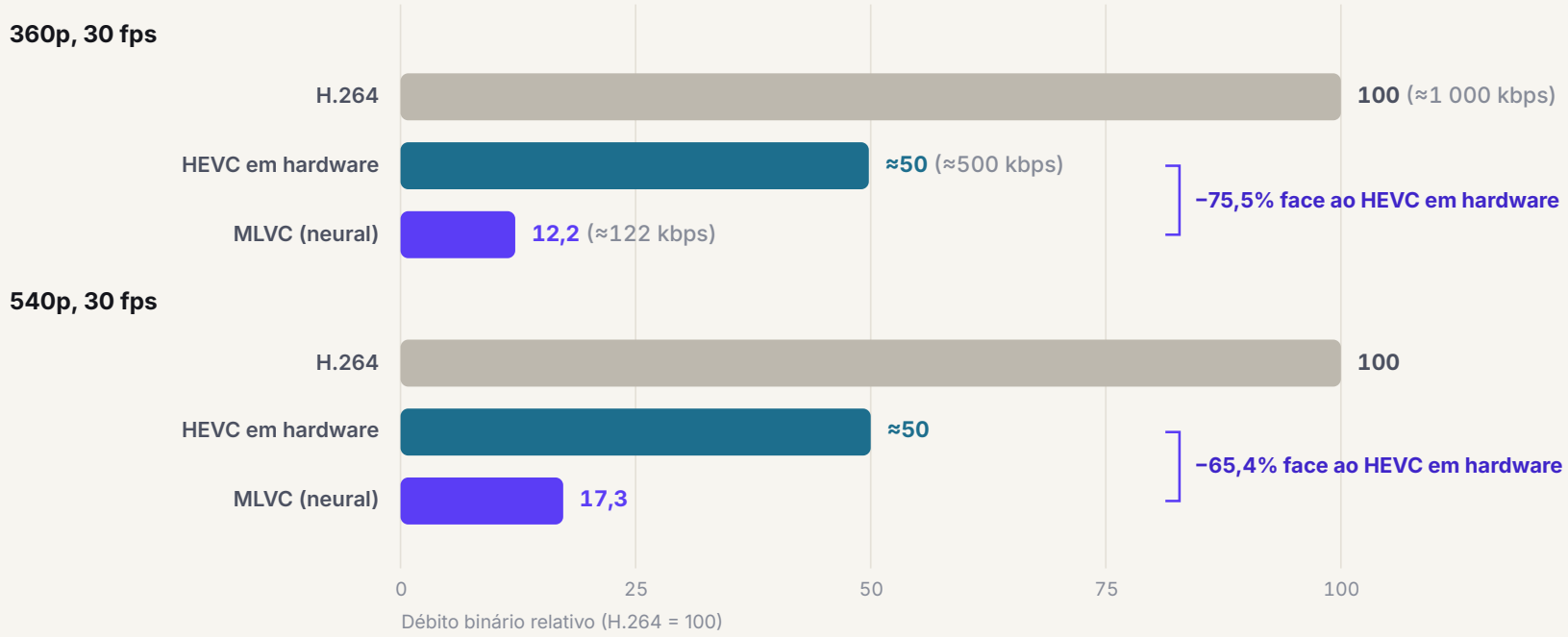
  Publicado em 2025: 12,5–27,9% face ao VVC intra, com 7 a 216 mil MAC por píxel. Não encontrei nenhum produto que o use por predefinição.

- **Web**

  O Chrome 155 ativou o JPEG XL por predefinição a 6 de outubro de 2026, com um descodificador em Rust com segurança de memória.

> **Os formatos que têm de abrir em todo o lado continuam clássicos; a codificação neural ganha dentro da cadeia de uma empresa.**

É nas chamadas que os codecs neurais de ponta a ponta chegam ao mercado



O Satin funciona em todas as chamadas a dois do Teams e do Skype desde 2021. O MLVC, de código aberto sob licença MIT desde 24 de julho de 2026, está a ser implementado nas chamadas ponto a ponto do Teams até 540p30, com no máximo metade da NPU **alegação do fornecedor**.

Nem todas as plataformas optaram pelo neural, e as razões são instrutivas

DENTRO DE UMA NORMA, OPCIONAL

Opus 1.5 e 1.6

Melhoria neural, ocultação de perda de pacotes, redundância profunda e extensão de largura de banda, tudo desativado por predefinição e totalmente compatível com a norma de 2012.

CLÁSSICO, POR OPÇÃO

Meta MLow (2024)

Mais de 20% das chamadas da Meta correm em telemóveis ARMv7, e dezenas de milhões de chamadas de WhatsApp por dia em telemóveis com mais de dez anos. O MLow dá cerca do dobro da qualidade do Opus a 6 kbps com menos 10% de computação.

O orçamento de complexidade da base instalada é que decide. Os codecs neurais chegam primeiro onde os dispositivos são novos e um fornecedor controla as duas pontas.

A maior implantação a que ninguém chama codec

- **Codecs aprendidos em todo o lado**

  Cada modelo de voz e de vídeo lê e escreve os tokens ou latentes de um codec aprendido.

- **Não são formatos de distribuição**

  Cada modelo tem o seu espaço latente: sem controlo de débito, sem teste de conformidade, sem segundo descodificador.

- **Uma linha de custo**

  Para quem serve vídeo generativo, uma representação latente mais compacta reduz o custo de geração mais do que proporcionalmente análise.

DOIS EXEMPLOS

12,5 Hz · 1,1 kbps

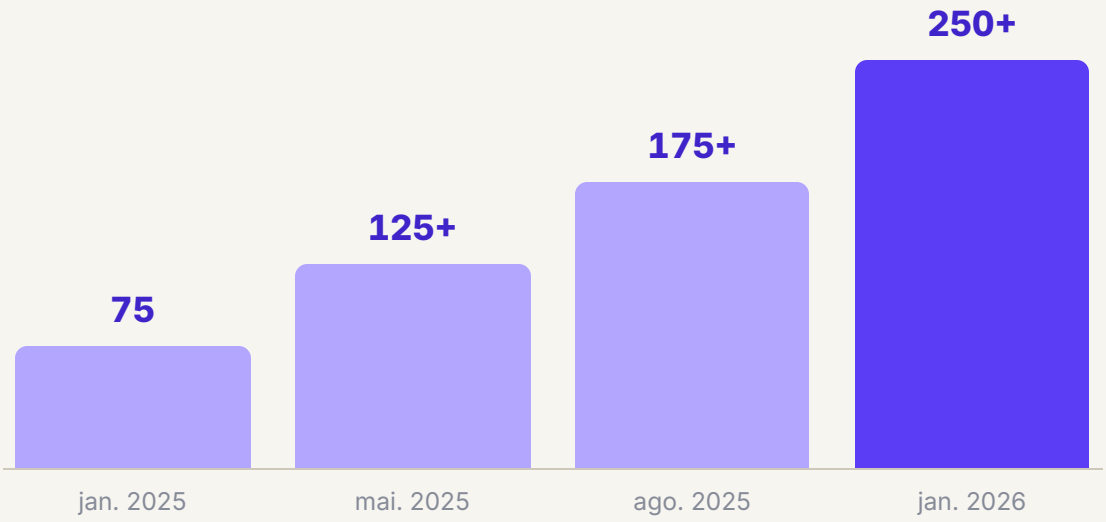
tokens de voz do Mimi, da Kyutai, abaixo de qualquer codec de telefonia

1:192

VAE de vídeo do LTX-Video: um token latente por 32×32 píxeis em 8 fotogramas

A reconstrução neural está generalizada; os recursos neurais estão a começar

DLSS 4 Multi Frame Generation: jogos e aplicações



Compressão neural de texturas

Cena de demonstração da NVIDIA, GTC 2026

Memória de texturas 6,5 GB → 0,97 GB

85% menos memória

Assassin's Creed Mirage, já lançado

Só recursos selecionados; face a formatos de blocos

≈30% mais compressão

Medidas diferentes: não diretamente comparáveis

> O DirectX Linear Algebra entrou em pré-visualização pública em abril de 2026: um só caminho de código para a descodificação neural em shaders, em qualquer GPU.

Satélites, drones e robôs: o caso mais adequado, um mercado pequeno

- **Porque se adequa**

  Um só dono controla as duas pontas, a largura de banda é escassa e o conteúdo é restrito o bastante para o treino.

- **China**

  O Anúncio n.º 49 de 2026, em vigor a 18 de novembro, revê os controlos de exportação de drones; uma das entradas usa a compressão semântica de imagem como limiar.

- **Estados Unidos**

  A FCC acrescentou os drones de fabrico estrangeiro e os seus componentes críticos à sua Covered List a 22 de dezembro de 2025.

> **Um nicho sólido para um especialista capaz de cumprir as regras de exportação, não um mercado de massas.**

Uma ordem de grandeza a mais em energia por píxel

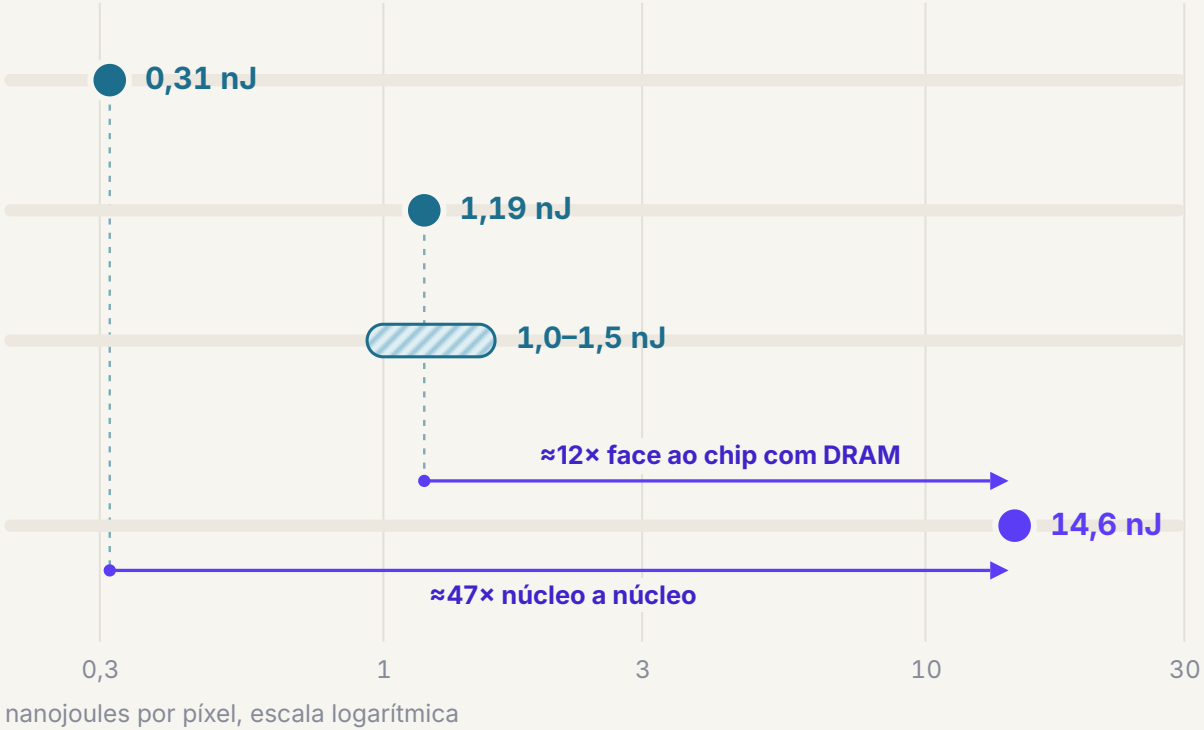
Chip descodificador HEVC, núcleo
40 nm, 4K30, medido

Chip descodificador HEVC, com DRAM
40 nm, 4K30, núcleo medido mais DRAM simulada

Descodificador VVC
28 nm, 4K60, projeção

Descodificador de vídeo neural (NVCA)
28 nm, 1080p25, só síntese; pesos e DRAM não contabilizados

A 1080p30, são cerca de 74 mW face a 908 mW.



A 1080p30: cerca de 74 mW para o chip HEVC com DRAM, face a aproximadamente 900 mW para o NVCA. Nó mais antigo no chip HEVC, pesos e DRAM não contabilizados no NVCA: a comparação favorece o projeto neural.

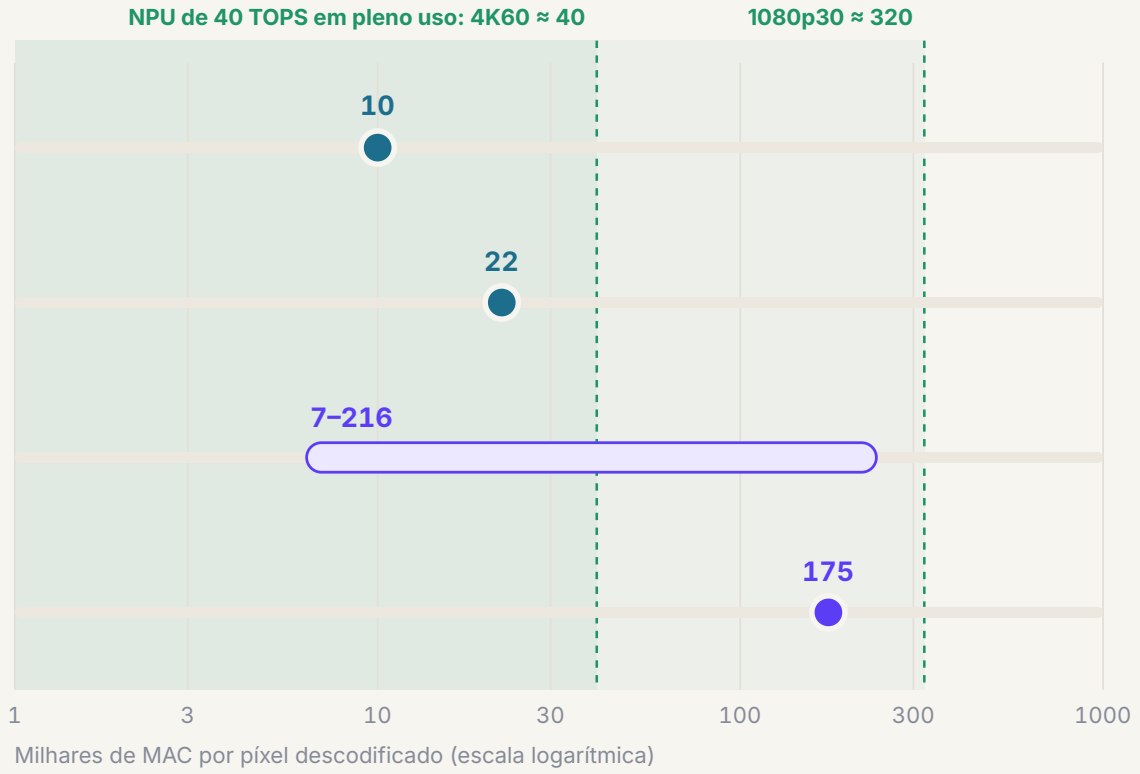
O que um descodificador neural pede a uma NPU

Filtro em ciclo neural, complexidade muito baixa
Exploração da JVET · ≈6% de poupança de bits
1,2 TOPS a 1080p30 · 10 a 4K60

Filtro em ciclo neural, complexidade baixa
Exploração da JVET · ≈8%
2,7 TOPS a 1080p30 · 22 a 4K60

Descodificadores JPEG AI, do base ao alto
12,5–27,9% face ao VVC intra

DCVC-RT, um descodificador de vídeo neural completo
21% face à referência do VVC, cenário do artigo
22 TOPS a 1080p30 · 174 a 4K60



O DCVC-RT precisa de cerca de 175 mil MAC por píxel: cerca de 22 TOPS a 1080p30 e 174 a 4K60, com utilização plena **estimativa**. Uma NPU de 40 TOPS em pleno uso comporta cerca de 40 mil a 4K60.

O tráfego de memória e a exatidão ao bit são os limites rígidos

A DRAM DOMINA

Memória antes dos MAC

No chip HEVC do MIT, a DRAM é cerca de três quartos da energia do sistema: 0,88 de 1,19 nJ por píxel. Os descodificadores neurais mantêm grandes mapas de características entre fotogramas; os autores do DCVC-RT viram que a E/S de memória, não a aritmética, limitava a velocidade.

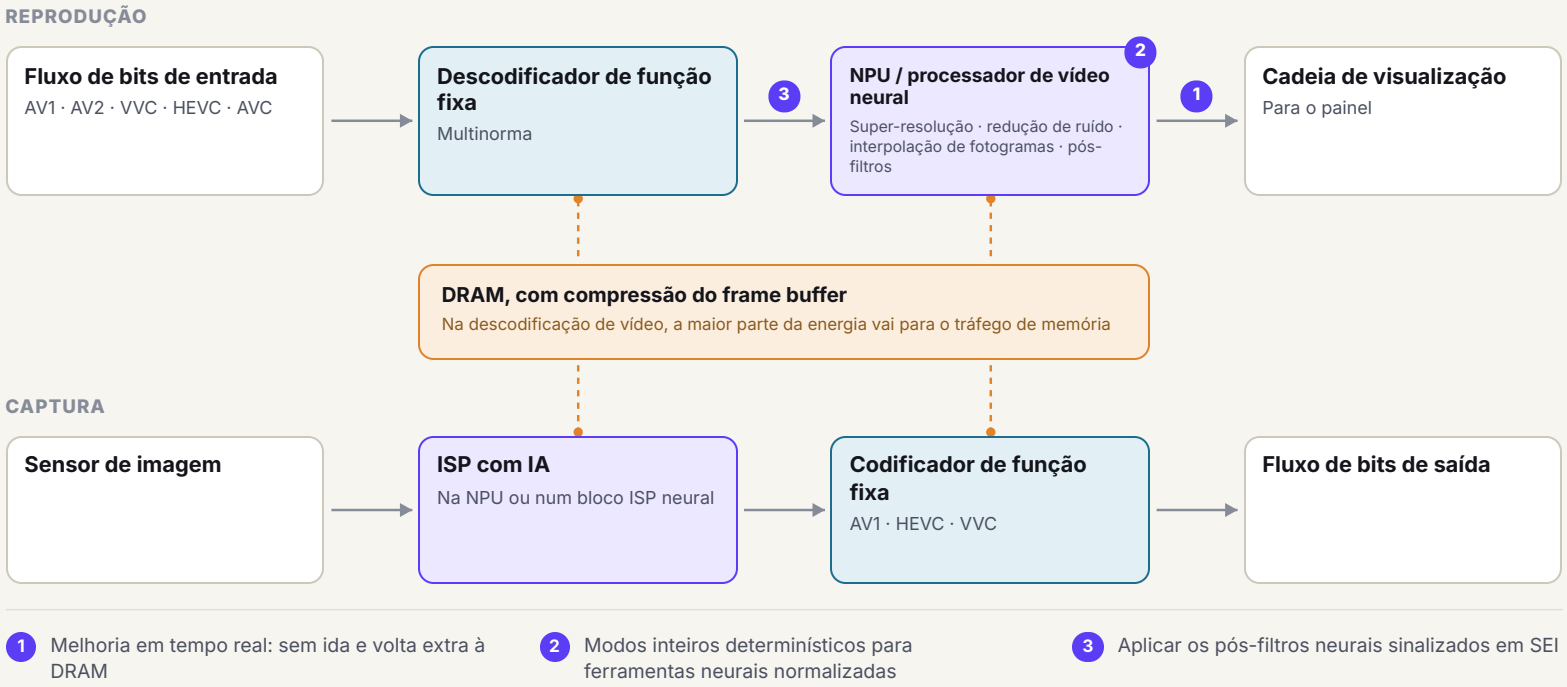
EXATIDÃO

Cada descodificador tem de calcular os mesmos bits

A descodificação entrópica falha se dois dispositivos calcularem uma probabilidade de forma diferente. Soluções: modelos inteiros (DCVC-RT), um hyperprior de escala partilhada (MLVC), inferência inteira de 16 bits (SADL da JVET), pontos de conformidade (JPEG AI).

Para uma equipa de chips, é de novo a disciplina das transformadas inteiras do HEVC: a norma tem de definir a aritmética com exatidão.

A divisão que chega ao mercado: um codec de função fixa mais uma NPU



> Canghai da Tencent: mais de 100 000 chips de primeira geração implantados; a segunda geração acrescenta H.266 e super-resolução no chip. A Chips&Media e a Allegro DVT vendem IP de codec e IP de vídeo neural lado a lado.

O que eu construiria, por perfil

Equipas de SoC e chips de TV

- Descodificadores de função fixa para cada norma a reproduzir; AV2 desde já
- Uma NPU dimensionada para melhoria à resolução do ecrã, sem ida e volta extra à DRAM
- SEI de pós-filtro neural e modos inteiros determinísticos
- O H.267 como início de projeto em 2029 ou depois

Fornecedores de IP

- Descodificadores AV2 e VVC, já
- Aceleradores de pós-filtros e compressão do frame buffer
- Ferramentas de conformidade para processamento neural

Plataformas de vídeo

- O pré-processamento neural no codificador compensa primeiro
- Super-resolução no cliente onde os dispositivos o permitam
- Codecs neurais de ponta a ponta só em circuitos fechados que controle

Onde as ferramentas neurais se tornam licenciáveis

Norma	Conteúdo neural	Situação, novembro de 2026
H.267 (nome provisório)	Em aberto: ferramentas clássicas (ECM) e neurais são ambas candidatas	Respostas avaliadas em janeiro de 2027; versão 1 no final de 2029
VVC e HEVC, com VSEI	Pós-filtro neural descrito em SEI; opcional para os descodificadores	Em vigor; alargado em outubro de 2025
AV2	Clássico, isento de royalties	Versão 1.0.0 a 28 de maio de 2026; YouTube prevê emissões na Web no início de 2027
JPEG AI	Codec de imagem aprendido de ponta a ponta	Parte 1 de 2025; sem suporte por predefinição encontrado em produtos
IEEE 1857.11	Codificação de imagem neural concebida para NPU	Publicada em dezembro de 2024
Video Coding for Machines	Codificação para análise por máquinas e não para visualização	Projeto de norma internacional no início de 2026
Extensões de ML do Opus	Opcionais e retrocompatíveis	No libopus 1.5 e 1.6

645

famílias de patentes pós-VVC detidas pela ByteDance (GreyB, 2025)

0,50 USD · 1,00–1,50 USD

tarifa do agrupamento VVC por telemóvel e por TV 4K, Região 1

0 USD

de royalties dos membros da AOMedia pelo AV1 e AV2

QUEM GANHA DINHEIRO

O dinheiro segue as normas e as plataformas

JÁ A FATURAR

Plataformas

Ficam com as poupanças

Redutor de resolução neural da Netflix; super-resolução do YouTube e da Meta; a Microsoft abriu o código do MLVC sob licença MIT

JÁ A FATURAR

Fabricantes de dispositivos e chips

Vendem diferenciação

ISP de câmaras; processadores de ampliação para televisores; chips de codec Canghai, próprios da Tencent

JÁ A FATURAR

Fornecedores de IP

Licenciam cada nova norma

Clientes para descodificadores AV2 na Chips&Media, Allegro DVT e VeriSilicon nas três semanas seguintes à especificação

JÁ A FATURAR

Fabricantes de GPU

Criam dependência da plataforma

DLSS 4 em 250+ jogos e aplicações; SDK de texturas neurais em beta

POR DISPOSITIVO

Titulares e agrupamentos de patentes

Cobram por dispositivo

VVC: 0,50 USD por telemóvel, 1,00–1,50 USD por TV 4K; ferramentas do H.267 após 2029

SAÍDA POR AQUISIÇÃO

Startups de codecs neurais

Adquiridas

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

> As três startups de codecs neurais mais conhecidas, WaveOne, iSIZE e Deep Render, foram compradas pela Apple, pela Sony e pela InterDigital.

ONDE UMA EMPRESA PODE APOSTAR

Seis apostas e o risco de cada uma

Aposta	Quem compra	Porquê agora	Principal risco
Pré-processamento neural no codificador	Plataformas de streaming e de vídeo de utilizadores, radiodifusores	Sem mudar dispositivos; provado na Netflix	As grandes plataformas fazem-no em casa
IP e afinação para melhoria no ecrã	Fabricantes de chips para TV, boxes e telemóveis	As NPU já estão nos dispositivos	Banalização; equipas internas
Codecs neurais em circuito fechado	Fornecedores de comunicações empresariais, industriais e de telepresença	O MLVC prova que funciona, e o seu código é aberto	É preciso controlar as duas pontas; fragmentação das NPU
Silício ou IP de codec mais NPU	Fabricantes de SoC	O ciclo do AV2 está a começar	Incumbentes fortes
Avaliação, conformidade e inferência determinística	Quem lança ferramentas neurais	A exatidão ao bit é um requisito estrito	Mercado pequeno, salvo em pacote
Contribuições normativas para o H.267	Licenciantes, grandes plataformas	As respostas são avaliadas em janeiro de 2027	Longo, caro, incerto

Desvalorize uma alegação até ela indicar quatro coisas

Alegação	O que diz a evidência
«Os codecs neurais já batem o VVC.»	Em laboratório. O DCVC-RT poupa 21% face ao codificador de referência do VVC; o seu descodificador custa cerca de 175 mil MAC por píxel.
«Os codecs de IA vão substituir o H.26x e o AV1 dentro de poucos anos.»	Não na reprodução: versão 1 do H.267 no final de 2029, chips por volta de 2031 estimativa. O codec que está a chegar agora, o AV2, é clássico.
«O MLVC reduz a largura de banda em 88%.»	Face ao H.264 a 360p. Face ao HEVC em hardware: 75,5% a 360p e 65,4% a 540p, no conjunto de testes da própria Microsoft.
«As texturas neurais reduzem a memória de vídeo em 85–96%.»	Em cenas de demonstração. O jogo já lançado ganhou cerca de 30%, e só em recursos selecionados.
«O JPEG AI é uma norma ISO, por isso câmaras e navegadores vão adotá-lo.»	Nenhum suporte por predefinição encontrado em produtos. A Web adotou o JPEG XL, um formato clássico.

Pergunte pela âncora, pela métrica, pela resolução e pelo custo de descodificação. Uma alegação que não indica as quatro é um resultado de investigação, não um produto.

O que acompanhar, e questões em aberto

MARCOS

Janeiro de 2027	Avaliação das respostas ao H.267: quantas dependem de ferramentas neurais, a que custo por píxel
Início de 2027	Primeiras emissões AV2 do YouTube na Web; descodificadores AV2 em SoC anunciados
Segundo semestre de 2027	Primeiro modelo de teste do H.267
2027	O MLVC além das chamadas ponto a ponto do Teams
2027–2028	Um primeiro produto que use o JPEG AI por predefinição
Final de 2029	A versão 1 do H.267 inicia o ciclo de silício

QUESTÕES EM ABERTO

01 Exatidão em escala: descodificação inteira exata ao bit nas NPU de vários fornecedores, com que perda de compressão?

02 Memória antes dos MAC: o tráfego de DRAM por píxel do pós-processamento neural

03 O custo por píxel a partir do qual uma ferramenta neural deixa de compensar no H.267

04 Que circuitos fechados além das chamadas adotam primeiro codecs neurais de ponta a ponta

05 Como medir a fidelidade quando um descodificador gera detalhe

São áreas que podemos explorar juntos, através de intercâmbio técnico e de colaboração técnica.

i@tanshuai.com · tanshuai.com