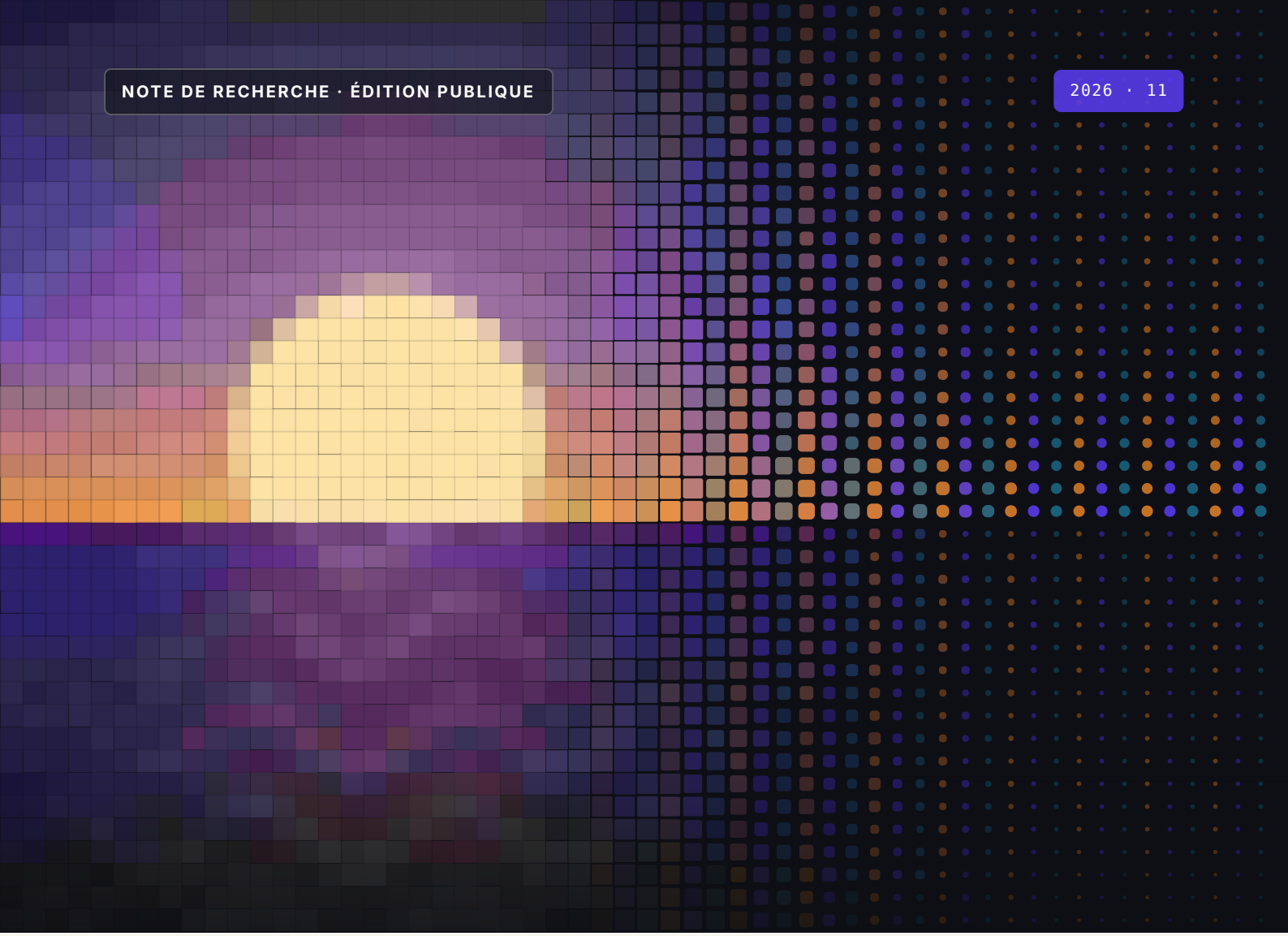


NOTE DE RECHERCHE · ÉDITION PUBLIQUE

2026 · 11

COMPRESSION APPRISE · LECTURE × SILICIUM

Le codage neuronal en 2026

Il se déploie autour du codec, pas à sa place

Lecture vidéo · Photographie · Appels · IA générative · Jeux · Silicium

LE VERDICT

Les réseaux de neurones interviennent déjà aux deux extrémités des codecs qui acheminent la vidéo mondiale. Les codecs neuronaux de bout en bout ne sont déployés qu'en boucle fermée, et les écrans grand public attendent H.267.

Tan Shuai

Technologue et fondateur · Chercheur et développeur en IA

tanshuai.com · i@tanshuai.com

Données au 12 novembre 2026

Montants en dollars américains

SOMMAIRE

Sommaire



◆ **Synthèse** — Déjà généralisé, mais autour du codec — 04

01 **Périmètre et méthode** — Ce que « codage neuronal » veut dire ici, et comment je l'ai examiné — 07

02 **Lecture et streaming** — Lecture et streaming : neuronal aux deux extrémités, normalisé au milieu — 09

03 **Photographie et images** — Photographie et images : neuronal dans l'appareil photo, classique sur le Web — 12

04 **Appels et réunions** — Appels et réunions : là où les codecs neuronaux de bout en bout sont réellement déployés — 14

05 **IA générative** — IA générative : le plus grand déploiement, que personne n'appelle un codec — 16

06 **Jeux et graphisme** — Jeux et graphisme : la reconstruction neuronale est généralisée, les assets neuronaux démarrent — 17

07 **Liaisons spécialisées** — Liaisons spécialisées : satellites, drones et robots — 19

08 **Silicium** — Le point de vue du silicium : énergie, calcul, mémoire et exactitude — 20

09 **Normes et brevets** — Normes et brevets : là où les outils neuronaux passent sous licence — 25

10 **Qui gagne de l'argent** — Qui gagne de l'argent, et où une entreprise peut se positionner — 26

11 **Affirmations et faits** — Les affirmations que vous entendrez, et ce que disent les faits — 28

12 **Jalons** — À surveiller, de 2027 à 2030 — 29

13 **Questions ouvertes** — 30

◆ **À propos de l'auteur** — Tan Shuai — 31

◆ **Sources** — 32

FIGURES

Figures

FIG. 01	Où le codage neuronal est déployé en 2026	06
FIG. 02	Trois façons pour un réseau de neurones d'entrer dans un codec	08
FIG. 03	D'une norme finalisée aux écrans grand public, il faut près d'une décennie	11
FIG. 04	Dans les appels, un codec neuronal économise 65–76 % de débit face au HEVC matériel	15
FIG. 05	Quand une entreprise contrôle toute la pile, la reconstruction neuronale se diffuse vite	18
FIG. 06	Énergie par pixel décodé : les décodeurs à fonction fixe ont un ordre de grandeur d'avance	21
FIG. 07	Un outil neuronal coûte des dizaines de milliers de MAC par pixel ; un décodeur neuronal complet, des centaines de milliers	22
FIG. 08	Le partage des tâches dans les produits : du matériel à fonction fixe pour le flux binaire, un moteur neuronal pour les pixels	23
FIG. 09	Qui gagne de l'argent avec le codage neuronal en 2026	26

Étiquettes dans le texte

- **estimation** — calcul ou déduction de l'auteur à partir de données citées
- **selon le fournisseur** — publié par l'entreprise qui vend le produit, sans vérification indépendante
- **analyse** — interprétation des faits par l'auteur

Ceci est une traduction de l'édition anglaise, qui fait foi en cas de divergence.

Données au 12 novembre 2026 · Montants en dollars américains

SYNTHÈSE

Déjà généralisé, mais autour du codec

LE VERDICT

Le codage neuronal est déjà généralisé en 2026, mais il opère autour des codecs normalisés, pas à leur place.

≈ 75 %
du trafic de données mobiles est de la vidéo, fin 2025 (Ericsson)

1,26 milliard
de smartphones livrés en 2025, qui décodent tous la vidéo avec du matériel à fonction fixe

≈ 12×
l'énergie par pixel d'une puce de décodage HEVC mesurée avec sa DRAM, pour la seule conception publiée de puce de décodage neuronal

≈ 22 TOPS
de calcul soutenu pour un décodeur neuronal de classe DCVC-RT en 1080p30 estimation

75,5 %
de débit en moins que le HEVC matériel pour MLVC en 360p dans les appels, sur le propre jeu de test de Microsoft
selon le fournisseur

Fin 2029
version 1 de H.267, la première norme susceptible d'apporter des outils neuronaux à tous les écrans

Où il est déployé

Les réseaux de neurones fonctionnent désormais aux deux extrémités des codecs qui acheminent la quasi-totalité de la vidéo mondiale. À l'entrée, ils façonnent l'image : processeurs de signal d'image à IA dans les appareils photo des téléphones, et sous-échantillonneur neuronal dans la chaîne d'encodage de Netflix. À la sortie, ils la réparent : Super Resolution de YouTube sur les téléviseurs, super-résolution chez Meta, et processeurs TV qui revendiquent des centaines de réseaux de neurones. Entre les deux, le flux binaire reste en AV1, HEVC, AVC ou VVC, décodé par du matériel à fonction fixe. Aucun service de streaming n'envoie de format vidéo neuronal aux téléviseurs ou aux téléphones ; lorsqu'un décodeur vidéo neuronal fonctionne aujourd'hui sur un appareil, il fonctionne à l'intérieur de l'application d'une seule entreprise.

Boucles fermées

Les codecs neuronaux de bout en bout, dont l'encodeur et le décodeur sont tous deux appris, ne sont déployés qu'en boucle fermée, là où une seule entreprise contrôle les deux extrémités et peut se replier sur un codec normalisé. Microsoft fait fonctionner un codec audio à IA dans tous les appels Teams et Skype à deux participants depuis 2021, et, en juillet 2026, l'entreprise a publié en open source MLVC, un codec vidéo neuronal en cours de déploiement dans les appels Teams pair à pair : en 360p, il nécessite environ 75 % de débit en moins que le HEVC matériel à qualité notée égale. Chaque modèle de génération vocale et vidéo repose sur un codec appris, des tokeniseurs de parole à 12,5 Hz aux VAE vidéo au taux de compression de 1:192. La compression neuronale de textures a été livrée dans un jeu AAA. Ce sont des déploiements réels, et ils ont une condition en commun : personne d'autre n'a à décoder le flux.

Calendrier

La lecture de masse de flux binaires neuronaux attend une norme et un cycle silicium. La première norme susceptible d'embarquer des outils neuronaux, le successeur de VVC que j'appellerai H.267, a lancé son appel à propositions en juillet 2026, évaluera les réponses en janvier 2027 et vise une première version fin 2029. Il a fallu près de huit ans à l'AV1 pour passer de la spécification finalisée à 30 % des visionnages Netflix. Sur ce modèle, les premières puces H.267 arriveront vers 2031 et l'usage se généralisera au milieu des années 2030 estimation. En attendant, le prochain codec à atteindre les écrans est l'AV2, finalisé en mai 2026 sans aucun décodeur neuronal, et trois fournisseurs d'IP ont annoncé des clients pour leurs décodeurs AV2 en moins de trois semaines.

Silicium

La raison tient au silicium. La seule conception publiée de puce de décodage vidéo neuronal nécessite 14,6 nJ par pixel, soit environ 12 fois une puce de décodage HEVC mesurée, DRAM comprise, et environ 47 fois son seul cœur. Selon mon décompte à partir de son propre tableau par module, le codec de recherche temps réel de Microsoft, DCVC-RT, nécessite environ 175 000 multiplications-accumulations par pixel décodé, soit environ 22 TOPS de calcul soutenu en 1080p30 estimation : à peu près la moitié des 40 TOPS qui définissent un PC IA, consacrée à un seul flux vidéo. Un décodeur neuronal doit en outre produire des résultats identiques au bit près sur chaque appareil, faute de quoi son décodage entropique s'effondre. C'est pourquoi MLVC sacrifie de la compression pour décoder de façon identique sur tous les NPU, et fonctionne en 540p, pas en 4K.

Économie

Pour une entreprise, l'argent suit les normes et les plateformes, pas les codecs neuronaux autonomes. Les plateformes gardent les économies du pré- et du post-traitement neuronal. Les fournisseurs d'IP vendent des licences de décodeurs à fonction fixe pour chaque nouvelle norme, plus des NPU pour l'imagerie. Les pools de brevets perçoivent des redevances par appareil : 0,50 dollar par téléphone et 1,00–1,50 dollar par téléviseur 4K pour VVC aux États-Unis et sur les autres marchés de la Région 1. Les trois start-up de codecs neuronaux les plus connues, WaveOne, iSIZE et Deep Render, ont été rachetées par Apple, Sony et InterDigital. Les opportunités ouvertes se trouvent dans le pré- et le post-traitement neuronal, dans les produits en boucle fermée, dans le silicium qui associe un codec à fonction fixe et un NPU, et dans le travail d'évaluation et de normalisation entre les deux.

Question	Appréciation	Chiffres clés (sources dans les chapitres)
① **Où le codage neuronal est-il déployé à grande échelle aujourd'hui ?**	Autour du codec : ISP des appareils photo, sous-échantillonnage côté encodage, super-résolution à la lecture. De bout en bout uniquement en boucle fermée : appels, assets de jeu, tokeniseurs d'IA générative	Vidéo ≈ 75 % des données mobiles ; 1,26 milliard de téléphones et ≈ 200 millions de téléviseurs livrés en 2025, qui décodent tous avec du matériel à fonction fixe ; DLSS 4 dans plus de 250 jeux et applications
② **Les codecs neuronaux remplaceront-ils AV1, HEVC et VVC pour la lecture ? Quand ?**	Pas avant une norme et un cycle d'appareils. H.267 est le premier vecteur réaliste, et il mêlera probablement outils classiques et neuronaux	Appel à propositions H.267 en juillet 2026, évaluation en janvier 2027, version 1 fin 2029 ; généralisation au milieu des années 2030 estimation
③ **Que coûte-t-il en silicium ?**	Un ordre de grandeur de plus en énergie par pixel que le décodage à fonction fixe actuel, plus le trafic mémoire et une exigence d'exactitude au bit près	14,6 nJ/pixel (neuronal) contre 1,19 (HEVC, avec DRAM) ; ≈ 175 kMAC par pixel décodé ≈ 22 TOPS en 1080p30 estimation

Question	Appréciation	Chiffres clés (sources dans les chapitres)
④ **Que permettent les normes et les brevets ?**	Les post-filtres neuronaux peuvent déjà être signalés dans les flux VVC ; JPEG AI est une norme ISO ; H.267 décidera quels outils neuronaux passeront sous licence	ECM (classique) ≈ 28 % de mieux que VVC ; filtres en boucle neuronaux 6–14 % pour 10–22+ kMAC/pixel ; ByteDance : 645 familles de brevets dans les travaux post-VVC
⑤ **Qui gagne de l'argent, et où une entreprise peut-elle se positionner ?**	Plateformes, fournisseurs d'IP, détenteurs de brevets. Les start-up de codecs purement neuronaux ont été rachetées. Se positionner sur le pré- et post-traitement, les boucles fermées, le silicium codec plus NPU, l'évaluation	Pool VVC : 0,50 dollar par téléphone ; IP de décodage AV2 concédée sous licence quelques semaines après la spécification ; trois acquisitions, conditions non divulguées
⑥ **Quelles affirmations faut-il relativiser ?**	Les gains annoncés par rapport à d'anciens codecs, en basse résolution, sans coût de décodage, ou issus de démonstrations plutôt que de produits commercialisés	MLVC : 88 % de moins que H.264, mais 65–76 % de moins que le HEVC matériel ; textures : 85 % de VRAM économisée dans une démo, environ 30 % dans le jeu commercialisé

FIG. 01 **Où le codage neuronal est déployé en 2026**

Les réseaux de neurones fonctionnent déjà aux deux extrémités des codecs courants. Les codecs neuronaux de bout en bout ne sont déployés que là où une seule entreprise contrôle à la fois l'encodeur et le décodeur.

	Autour d'un codec normalisé prétraitement, amélioration, mise à l'échelle	Codec neuronal de bout en bout encodeur et décodeur appris	Dans une norme outils neuronaux dans un flux binaire ou une API normalisés
01 **Lecture et streaming**	● À GRANDE ÉCHELLE Sous-échantillonneur de Netflix ; super-résolution chez YouTube et Meta ; mise à l'échelle par IA des téléviseurs	● NON DÉPLOYÉ Aucun service de streaming n'envoie de format vidéo neuronal ; téléviseurs et téléphones décodent les codecs normalisés en matériel	● EN NORMALISATION La version 1 de H.267 est attendue fin 2029 ; des outils neuronaux figurent parmi les candidats
02 **Photographie et imagerie**	● À GRANDE ÉCHELLE ISP IA dans les téléphones haut de gamme ; un ISP entièrement neuronal présenté comme IP au CES 2026	● DÉMOS, PILOTES JPEG AI est une norme ISO : démos sur téléphone, aucun produit	● NON DÉPLOYÉ Le format qui progresse est classique : JPEG XL activé par défaut dans Chrome 155
03 **Appels et réunions**	● À GRANDE ÉCHELLE Super-résolution côté récepteur et suppression neuronale du bruit	● À GRANDE ÉCHELLE Audio neuronal dans les appels Teams à deux depuis 2021 ; vidéo MLVC en cours de déploiement dans Teams	● DÉPLOIEMENT LIMITÉ Opus 1.5 ajoute des outils ML optionnels et rétrocompatibles
04 **IA générative**	—	● À GRANDE ÉCHELLE Chaque modèle vidéo et vocal utilise un tokeniseur appris, par ex. un VAE vidéo 1:192	—
05 **Jeux et graphisme**	● À GRANDE ÉCHELLE Mise à l'échelle neuronale et génération d'images : DLSS 4 seul dans plus de 250 jeux et applications	● DÉPLOIEMENT LIMITÉ Textures neuronales livrées dans un jeu AAA ; le SDK de NVIDIA est en bêta	● EN NORMALISATION DirectX ajoute des opérations matricielles pour les shaders (préversion)
06 **Liaisons spécialisées : satellites, robots**	● DÉMOS, PILOTES Restauration côté serveur de flux à faible débit	● DÉMOS, PILOTES Codecs neuronaux résistants aux pertes, en recherche et en pilote	—

Source: Documents d'entreprises, de normalisation et de développeurs cités dans le rapport ; situation au 12 novembre 2026. Les lignes sont classées selon le nombre de personnes que touche chaque domaine.

PÉRIMÈTRE ET MÉTHODE

01 Ce que « codage neuronal » veut dire ici, et comment je l'ai examiné

Trois façons pour un réseau de neurones d'entrer dans un codec

Le « codage neuronal » recouvre trois choses différentes, et la plupart des débats à son sujet viennent de leur confusion.

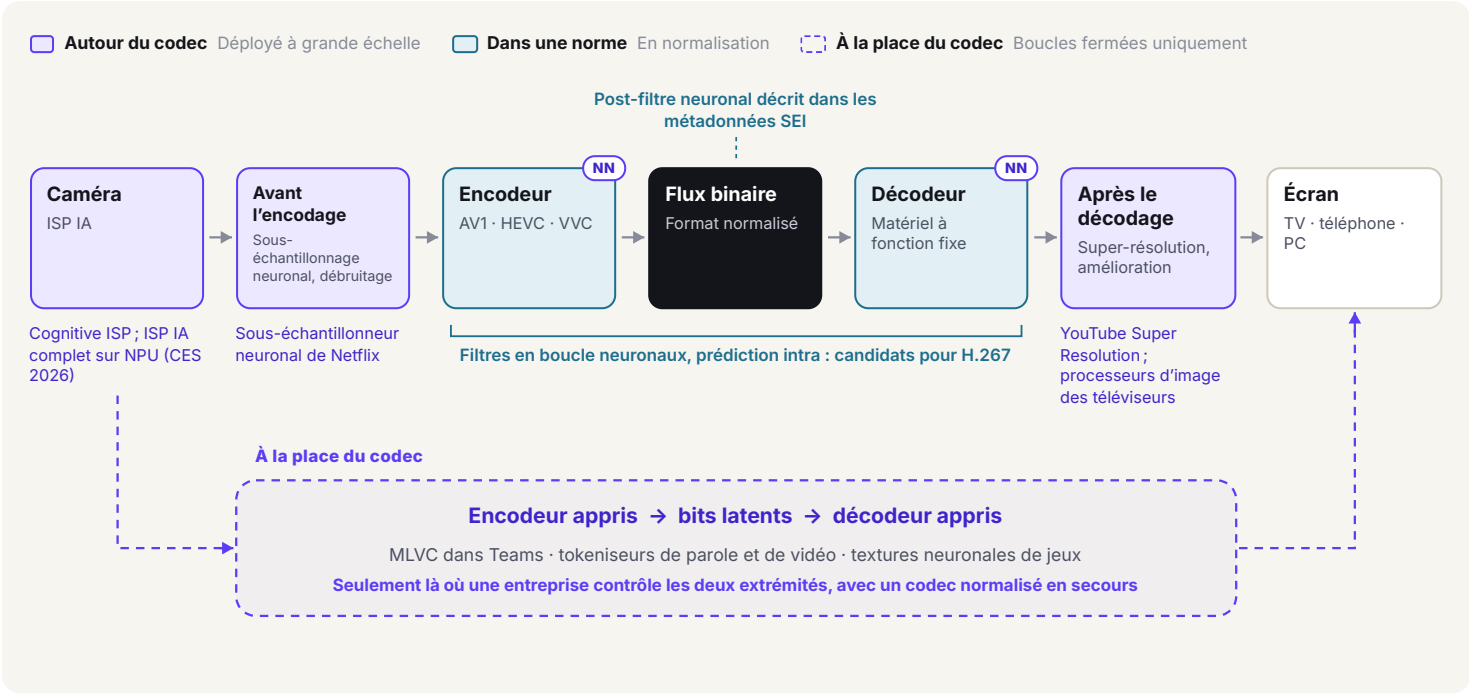
01 **Autour du codec.** Un réseau traite les pixels avant l'encodage ou après le décodage, et le flux binaire reste normalisé. Exemples : un processeur de signal d'image (ISP) à IA dans un appareil photo, le sous-échantillonneur neuronal que Netflix place avant son encodeur, la super-résolution sur un téléviseur. N'importe quel appareil peut lire le flux ; le réseau ne modifie que ce qui entre ou ce qui sort.

02 **Dans une norme.** Un réseau devient un outil d'un codec normalisé, par exemple un filtre en boucle neuronal ou un prédicteur intra neuronal que chaque décodeur doit exécuter ; il doit donc être spécifié exactement, généralement en arithmétique entière. Une forme plus légère est un post-filtre neuronal que le flux binaire décrit dans des métadonnées et que les décodeurs peuvent choisir d'appliquer.

03 **À la place du codec.** L'encodeur et le décodeur sont tous deux appris de bout en bout. Le flux binaire est un ensemble de valeurs latentes quantifiées que seul le réseau correspondant peut décoder. Exemples : MLVC de Microsoft dans Teams, les tokeniseurs de parole des modèles vocaux, les textures neuronales d'un jeu.

FIG. 02 **Trois façons pour un réseau de neurones d'entrer dans un codec**

L'essentiel de ce qui est déployé en 2026 opère autour d'un codec normalisé. Des outils neuronaux au sein d'une norme sont en cours de spécification. Les codecs neuronaux de bout en bout ne remplacent le codec que là où une seule entreprise contrôle les deux extrémités.



Source: Synthèse de l'auteur à partir des sources de ce rapport ; situation au 12 novembre 2026.

La première forme ne demande la permission de personne. La deuxième exige une norme. La troisième exige de contrôler les deux extrémités. C'est l'ordre dans lequel elles atteignent le public, et cela explique l'essentiel de ce qui suit.

Un classement fondé sur le monde réel, pas sur l'agenda de la recherche

Les articles de recherche étudient surtout la troisième forme, parce que c'est là que se trouvent les records de compression. Ce rapport classe les domaines selon le nombre de personnes qu'ils touchent, parce que c'est ce qui décide où vont l'argent et l'ingénierie : la lecture vidéo et le streaming d'abord, puis la photographie, les appels, l'IA générative, les jeux, et les liaisons spécialisées (satellites, drones, robots) en dernier. La vidéo représente environ 75 % de tout le trafic de données mobiles, selon le décompte d'Ericsson à la fin de 2025. En 2025, l'industrie a livré environ 1,26 milliard de smartphones (IDC) et environ 196–209 millions de téléviseurs (estimations de TrendForce et d'Omdia). Chacun d'eux décode la vidéo avec du matériel à fonction fixe.

Ma méthode de travail

Il s'agit d'une recherche documentaire, menée comme j'examinerais une puce ou un plan produit. J'ai lu des documents de normalisation et des comptes rendus de réunion (JVET, MPEG, JPEG, AOMedia, IETF), des blogs d'ingénierie et des notes de version d'entreprises, des articles sur des puces, des grilles tarifaires de pools de brevets, des textes sur le contrôle des exportations et la presse. Chaque chiffre est daté. Lorsque je calcule moi-même un chiffre, par exemple le calcul par pixel à partir des tableaux d'un article, je le dis et je le marque estimation. Les affirmations que je n'ai pas pu vérifier auprès d'une source primaire sont marquées selon le fournisseur. Les montants sont en dollars américains. La situation est arrêtée au 12 novembre 2026.

Je compare ce qui est comparable quand je le peux, et je le signale quand je ne le peux pas. Les gains de compression ne veulent pas dire grand-chose sans quatre précisions : la référence (gain par rapport à quoi), la mé-

trique (PSNR, VMAF ou un panel de spectateurs), le contenu et la résolution, et le coût du décodage. L'essentiel des désaccords sur les codecs neuronaux disparaît dès que ces quatre éléments sont précisés.

CHAPITRE 02

LECTURE ET STREAMING

02 Lecture et streaming : neuronal aux deux extrémités, normalisé au milieu

Ce que reçoivent les spectateurs reste un flux binaire normalisé

Les plus grands services vidéo ont achevé une transition de codec et entament la suivante, et aucune des deux n'implique de décodeur neuronal. En décembre 2025, Netflix a indiqué que l'AV1 représentait environ 30 % de ses visionnages, ce qui en faisait son deuxième codec le plus utilisé après l'AVC. L'AV1 n'a atteint les téléviseurs que lorsque les décodeurs matériels y sont arrivés : Netflix a ajouté l'AV1 à sa certification d'appareils en 2019, l'a diffusé vers les téléviseurs à partir de fin 2021, et indique que 88 % des appareils grand écran soumis à certification entre 2021 et 2025 le prenaient en charge, presque tous depuis 2023. YouTube a déclaré qu'au 19 mars 2026, plus de 90 % de son catalogue, pondéré par le temps de visionnage, était disponible en AV1.

L'AV2, successeur de l'AV1 issu de l'Alliance for Open Media, a atteint la version 1.0.0 le 28 mai 2026, avec une compression environ 30 % meilleure que celle de l'AV1 selon les mesures objectives d'AOMedia. YouTube prévoit des flux AV2 sur le Web « dès début 2027 ». Le silicium a suivi aussitôt : Chips&Media a achevé un décodeur matériel AV2 en mai et l'a concédé sous licence en juin à une grande entreprise technologique nord-américaine non nommée ; Allegro DVT avait livré son IP de décodeur AV2 à trois clients au 16 juin ; VeriSilicon avait fourni à plusieurs clients une IP vidéo compatible AV2 au 10 juin. L'AV2 est un codec classique : ses gains viennent d'un meilleur partitionnement, d'une meilleure prédiction, de meilleures transformées et de meilleurs filtres en boucle (dont un filtre de détail guidé « piloté par les données »), pas d'un réseau de bout en bout.

Le VVC, la norme ITU/ISO actuelle, a trouvé son premier débouché en télédiffusion au Brésil, dont le système DTV+ (TV 3.0) est entré en exploitation commerciale en juin 2026, à temps pour la Coupe du monde. En juillet 2026, aucun téléviseur vendu au Brésil n'intégrait de récepteur DTV+ ; les premiers sont attendus entre fin 2026 et début 2027.

À l'encodeur : le sous-échantillonneur neuronal de Netflix

L'exemple le plus net de codage neuronal dans la lecture vidéo fonctionne depuis 2022 et ne change rien à ce que doit faire l'appareil du spectateur. Netflix encode chaque titre en plusieurs résolutions, et l'étape de sous-échantillonnage qui produit le 1080p, le 720p et le 540p à partir d'un master 4K utilisait traditionnellement un filtre fixe comme Lanczos. Netflix l'a remplacé par un petit réseau de neurones, un bloc de prétraitement suivi d'un bloc de redimensionnement, entraîné à améliorer la qualité perçue après une mise à l'échelle standard sur l'appareil. Netflix le décrit comme une solution prête à l'emploi : aucun autre changement côté encodage ni côté client, si bien que « des millions d'appareils compatibles avec le streaming Netflix en bénéficient automatiquement ». Dans les tests de préférence de Netflix, environ 77 % des spectateurs ont choisi le résultat du sous-échantillonneur neuronal **selon le fournisseur**. Pour que le coût reste acceptable à l'échelle de Netflix, « quelques couches de réseau de neurones ont suffi ».

C'est le schéma à retenir. Le réseau s'exécute une fois par titre dans le centre de données, où le calcul est bon marché et peut être planifié, et sa sortie est un flux AV1 ou HEVC ordinaire.

À l'écran : la super-résolution dans l'application et dans le téléviseur

À l'autre extrémité, les réseaux réparent ce que le codec et la source ont perdu. Le 29 octobre 2025, YouTube a annoncé Super Resolution pour son application TV : les vidéos mises en ligne en dessous de 1080p, de 240p à 720p, sont mises à l'échelle en HD par apprentissage automatique, avec la 4K comme objectif affiché. Les versions mises à l'échelle sont signalées, les spectateurs peuvent revenir à l'original et les créateurs peuvent refuser le traitement. Meta a décrit en octobre 2025 comment elle applique la super-résolution vidéo à la fois sur ses serveurs au moment de la mise en ligne, sur des processeurs x86 standard en partenariat avec Intel, et sur le client à la lecture, et a constaté que les deux se complètent. Les fabricants de téléviseurs font de même dans le silicium : le processeur 8K haut de gamme de Samsung, le NQ8 AI Gen3, revendique 768 réseaux de neurones pour la mise à l'échelle.

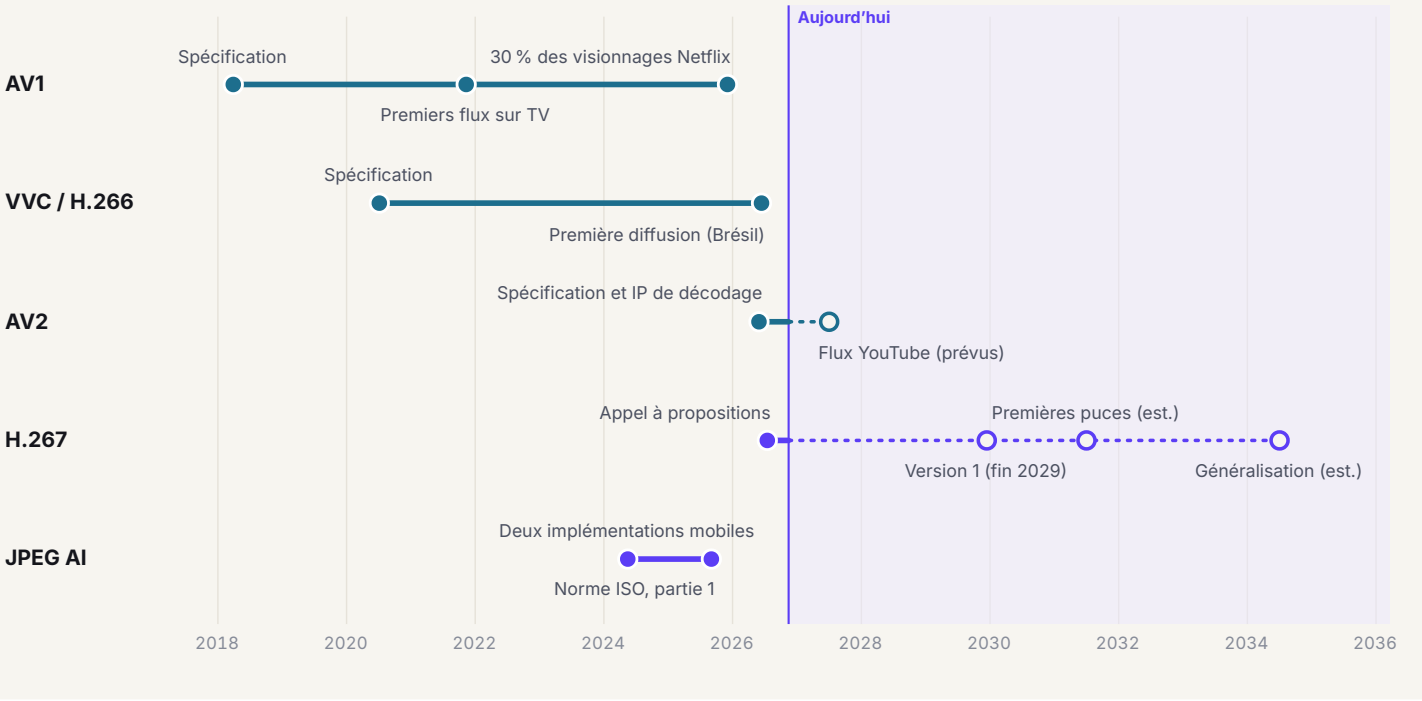
L'économie diffère selon l'endroit où tourne le réseau. Le traitement côté serveur coûte à la plateforme du calcul pour chaque vidéo traitée. Le traitement côté appareil ne coûte rien à la plateforme ; il utilise un NPU que l'acheteur a déjà payé. C'est pourquoi les gains les plus importants en 2026 viennent des deux.

Pourquoi aucun téléviseur ne décode de flux binaire neuronal

Un format de lecture doit pouvoir être décodé sur chaque appareil qui le lira un jour, et ces appareils durent des années. Le chemin entre une norme finalisée et les écrans grand public est donc lent, même pour les codecs classiques. La spécification de l'AV1 a été finalisée en mars 2018 ; le décodeur logiciel dav1d a suivi en quelques mois ; les téléviseurs l'ont reçu en streaming à partir de fin 2021 ; il a atteint environ 30 % des visionnages Netflix en décembre 2025, près de huit ans après la spécification. Le VVC a été finalisé en juillet 2020 et a atteint son premier service de télédiffusion en juin 2026.

FIG. 03 **D'une norme finalisée aux écrans grand public, il faut près d'une décennie**

La lecture exige un décodeur matériel dans l'appareil. Il a fallu près de huit ans à l'AV1 pour passer de la spécification finalisée à 30 % des visionnages Netflix ; la première norme susceptible d'embarquer des outils neuronaux est attendue fin 2029.



Source: Netflix Tech Blog (déc. 2025) ; AOMedia ; Allegro DVT ; TV Technology ; Fraunhofer IIS sur le DTV+ brésilien ; MPEG-155 ; comité JPEG ; ISO. Les marqueurs creux sont des objectifs ou des estimations de l'auteur.

Un flux binaire neuronal ajoute deux problèmes que les codecs classiques ont résolus depuis longtemps : l'énergie par pixel et le décodage exact au bit près. Tous deux sont traités au chapitre 8. Ensemble, ils font qu'aucun fabricant de téléviseurs ou de téléphones ne mettra de décodeur vidéo neuronal dans chaque appareil avant qu'une norme ne lui dise exactement quoi construire.

Ce qui pourrait changer la donne : H.267

La norme qui pourrait apporter des outils neuronaux à tous les écrans est le successeur de VVC, développé conjointement par l'ITU-T et l'ISO/IEC au sein du Joint Video Experts Team (JVET). Son nom n'est pas définitif ; je l'appellerai H.267, comme une bonne partie de l'industrie.

En octobre 2025, le JVET a évalué les réponses à un appel à preuves : cinq réponses de 16 entreprises et laboratoires de recherche, avec 20 configurations de codec testées. Le comité a trouvé des preuves d'une meilleure compression que VVC et a jugé « intéressantes » aussi bien la technologie conventionnelle améliorée que la technologie à base de réseaux de neurones. Les approches neuronales ont montré des gains perceptifs nets, en particulier sur les contenus 8K HDR, et des gains moindres sur les contenus de jeu vidéo. InterDigital, l'un des répondants, a soumis une conception fondée sur l'Enhanced Compression Model (ECM) conventionnel et une autre qui superposait des outils d'IA à VVC.

L'appel à propositions formel a été lancé lors de la réunion de juillet 2026 à Genève. Les principales soumissions étaient attendues fin octobre 2026, un complément est attendu à la mi-décembre, et les réponses seront évaluées lors de la réunion du JVET à Brisbane en janvier 2027. Le JVET prévoit un premier modèle de test au second semestre 2027 et une première version de la norme d'ici fin 2029, et qualifie le projet d'« effort probablement plus important que tout ce qui a jamais été fait pour lancer une nouvelle norme ».

La piste classique place la barre haut. L'ECM, construit avec un traitement du signal conventionnel, économise environ 28 % par rapport à VVC en configuration à accès aléatoire. L'exploration des réseaux de neurones menée par le JVET fait état de filtres en boucle qui économisent environ 6 % pour 10 000 multiplications-accumulations par pixel, 8 % pour 22 000, et 14 % à un point de fonctionnement de haute complexité. Ma lecture est que H.267 sera très probablement un cœur classique doté de quelques outils neuronaux là où les bits économisés justifient le coût au décodeur, et que l'évaluation de janvier 2027 montrera combien analyse.

Il existe déjà une passerelle. Les flux VVC et HEVC peuvent transporter des informations d'amélioration supplémentaires (SEI) qui décrivent un post-filtre neuronal à appliquer côté décodeur, et les nouvelles éditions de VVC et de sa norme SEI (VSEI), finalisées en octobre 2025, ont étendu ces messages. Un flux peut indiquer quel réseau exécuter après le décodage sans obliger chaque décodeur à l'exécuter.

Ce que cela signifie pour les entreprises de la lecture vidéo

Vendez dès maintenant ce qui fonctionne autour du codec : le prétraitement côté encodage pour les services, et les blocs de post-traitement et le réglage pour les puces de téléviseurs et de boîtiers TV. Ne bâtissez pas d'activité de lecture sur un flux binaire neuronal qui doit être décodé sur des appareils tiers avant les années 2030. Si vous fabriquez du silicium, les revenus à court terme sont les décodeurs AV2 et VVC ; la partie neuronale est une voie NPU d'amélioration placée à côté d'eux (chapitre 8).

CHAPITRE 03

PHOTOGRAPHIE ET IMAGES

03 Photographie et images : neuronal dans l'appareil photo, classique sur le Web

La chaîne de l'appareil photo est déjà neuronale

Chaque photo prise par un téléphone haut de gamme récent est passée par des réseaux de neurones avant d'être compressée. Le Snapdragon 8 Gen 2 de Qualcomm, présent dans les téléphones de 2023, a introduit un « Cognitive ISP » qui effectue une segmentation sémantique en temps réel pendant la capture photo et vidéo, en ajustant séparément les visages, les cheveux, le ciel et l'arrière-plan ; le Galaxy S23 Ultra de Samsung a été le premier téléphone à l'utiliser sur la caméra frontale. En janvier 2026, Chips&Media et Visionary.ai ont présenté au CES ce qu'elles appellent le premier ISP entièrement fondé sur l'IA : une chaîne d'imagerie de bout en bout qui fonctionne entièrement sur un NPU, le WAVE-N de Chips&Media, en virgule flottante 16 bits, sans ISP à fonction fixe, et qui peut être mise à jour à distance. « Les chaînes existantes greffent des blocs neuronaux sur un ISP à fonction fixe. Nous remplaçons entièrement l'ISP conventionnel », a déclaré le PDG de Visionary.ai. Il a été proposé aux fabricants pour évaluation après le salon.

Le résultat est ensuite enregistré en JPEG ou en HEIF, des formats classiques que n'importe quel appareil peut ouvrir. Le schéma est le même que dans la lecture vidéo : neuronal à l'intérieur de la chaîne d'une seule entreprise, normalisé à la frontière.

JPEG AI : une norme neuronale finalisée, encore sans produit

JPEG AI est la première norme de codage d'image apprise de l'ISO/IEC et de l'ITU-T (ISO/IEC 6048, ITU-T T.840). La partie 1, le système de codage de base, a été publiée en 2025. Le comité JPEG fait état de gains de compres-

sion de 12,5 % à 27,9 % par rapport au codage intra de VVC, sur une plage de complexité au décodeur allant de 7 000 à 216 000 multiplications-accumulations par pixel. Sa conception comprend deux encodeurs et trois décodeurs, entraînés conjointement, si bien qu'un récepteur peut n'implémenter qu'un seul décodeur et lire malgré tout chaque flux. En mai 2024, le comité faisait état de deux implémentations sur smartphone. En parallèle, l'IEEE a publié en décembre 2024 la norme IEEE 1857.11-2024, consacrée au codage d'image à base de réseaux de neurones et conçue pour les NPU.

En novembre 2026, je n'ai trouvé aucun appareil photo de téléphone, aucune application photo ni aucun navigateur qui écrive ou lise JPEG AI par défaut.

Ce que le Web a adopté à la place : JPEG XL

Le format d'image qui a vraiment progressé en 2026 est classique. Le 6 octobre 2026, Google a activé par défaut le décodage JPEG XL dans Chrome 155, avec un nouveau décodeur à sûreté mémoire écrit en Rust, quatre ans après en avoir retiré la prise en charge expérimentale. Safari décode JPEG XL depuis 2023, et Mozilla évolue dans le même sens. Les arguments de Google sont une meilleure compression que JPEG, la recompression sans perte des fichiers JPEG existants, le HDR et le décodage progressif.

La leçon se généralise. Un format qui doit s'ouvrir partout l'emporte avec un décodeur rapide et déterministe qui fonctionne sur n'importe quel processeur et dont la sûreté mémoire peut être garantie. Un décodeur d'image neuronal a besoin d'un NPU ou d'un GPU et d'une arithmétique exacte d'un appareil à l'autre. Le codage neuronal l'emporte à l'intérieur de la chaîne propre à un appareil et dans les systèmes où la même entreprise encode et décode, comme un service de photos dans le cloud qui ne transmet jamais le fichier compressé à personne d'autre.

Ce que cela signifie pour les entreprises de l'imagerie

L'argent se trouve dans la chaîne de capture : IP d'ISP IA, NPU dimensionnés pour l'imagerie, et le travail de réglage que les fabricants d'appareils photo achètent. Pour le stockage et la distribution, un codec d'image appris est aujourd'hui un outil pour systèmes fermés, pas un format sur lequel parier.

CHAPITRE 04

APPELS ET RÉUNIONS

04 Appels et réunions : là où les codecs neuronaux de bout en bout sont réellement déployés

L'audio est passé en premier

Microsoft a présenté Satin, un codec vocal fondé sur l'IA, en février 2021, en précisant qu'il était déjà utilisé pour tous les appels à deux participants dans Teams et Skype. Satin restitue la parole en super-large bande, avec une bande passante audio allant jusqu'à 16 kHz, à 6 kbps, là où l'ancien codec Silk de Skype était en bande étroite à ce débit. Le débit économisé sert à la redondance, ce qui permet aux appels de mieux résister aux pertes de paquets.

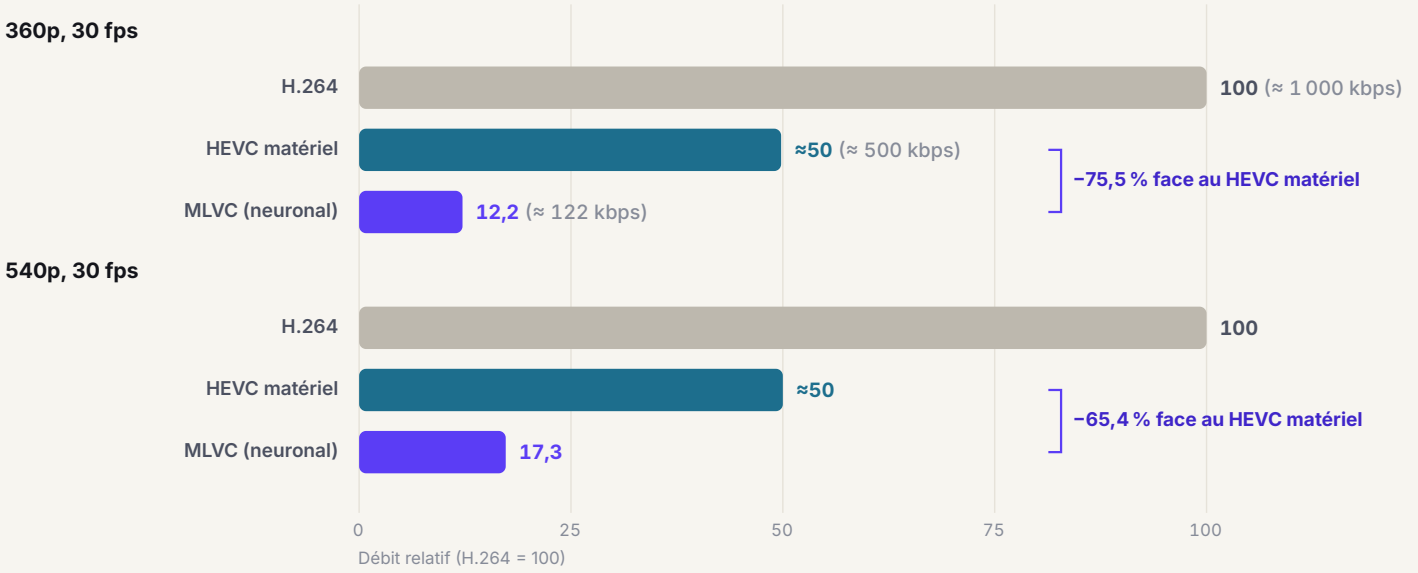
MLVC : un codec vidéo neuronal dans des appels en production

Le 24 juillet 2026, Microsoft a publié MLVC, son codec vidéo à apprentissage automatique, en open source sous licence MIT, avec le code du modèle, les poids entraînés, les scripts d'entraînement et les outils de conversion qui ciblent différents NPU. MLVC est la version produit de la famille DCVC que Microsoft Research publie depuis 2021. Microsoft indique déployer MLVC dans Teams pour les appels pair à pair, avec télémétrie en direct, tests A/B et les codecs conventionnels en secours.

MLVC fonctionne en 540p et 360p à 30 images par seconde et est conçu pour ne pas utiliser plus de la moitié du NPU, sur le Neural Engine d'Apple comme sur les NPU de Qualcomm et d'Intel, en répartissant le travail entre CPU et NPU. Lors de tests subjectifs selon ITU-T P.910 sur le propre Video Conferencing Dataset de Microsoft, il a nécessité 87,8 % de débit en moins que H.264 et 75,5 % de moins que le HEVC matériel en 360p, et 82,7 % et 65,4 % de moins en 540p selon le fournisseur. En 360p, cela représente environ 122 kbps, là où H.264 a besoin d'environ 1 Mbps.

FIG. 04 **Dans les appels, un codec neuronal économise 65–76 % de débit face au HEVC matériel**

Débit nécessaire à qualité notée égale sur le jeu de test de visioconférence de Microsoft, H.264 = 100. MLVC tourne sur les NPU d'ordinateurs portables et de téléphones jusqu'en 540p30, en utilisant au plus la moitié du NPU.



Source: Microsoft, publication de MLVC (24 juillet 2026) : tests subjectifs ITU-T P.910 sur son Video Conferencing Dataset. Barres HEVC déduites des économies annoncées. Données du fournisseur, pas des mesures indépendantes.

Deux éléments tempèrent le chiffre phare. Les 88 % sont mesurés par rapport à H.264 ; par rapport au HEVC matériel que contiennent déjà les mêmes ordinateurs portables, l'économie est de 65–76 %, ce qui reste considérable. Et MLVC a renoncé à une part de compression pour être déployable : une analyse indépendante note qu'un DCVC-RT entraîné sur un critère perceptif obtient 81,9 % sur la même mesure, là où MLVC obtient 75,5 %. Le prix d'un décodage identique sur du matériel que Microsoft ne possède pas a été d'environ six points.

Pourquoi les appels sont le terrain naturel

Un appel est une boucle fermée. Les deux terminaux exécutent la même application, négocient leurs capacités au début de l'appel et peuvent se replier sur un codec normalisé en cours d'appel. Les résolutions sont modestes, les réseaux souvent médiocres, et les ordinateurs portables des deux côtés sont désormais livrés avec des NPU, comme les téléphones depuis des années. Microsoft fixe à 40 TOPS de NPU le seuil de ses PC IA, et le Neural Engine de l'Apple M4 est annoncé à 38 TOPS. Toutes les conditions qui bloquent un codec neuronal dans la lecture vidéo sont absentes ici.

Les contre-exemples : Opus et MLow

Toutes les plateformes d'appel n'ont pas fait le même choix, et leurs raisons sont instructives.

Opus, le codec audio ouvert de WebRTC et de la plupart des applications d'appel, a ajouté l'apprentissage profond dans sa version 1.5 (mars 2024), sa première utilisation de réseaux pour traiter ou générer le signal lui-même. Mais il les a ajoutés sous forme d'options : une amélioration côté décodeur (LACE et NoLACE) qui ne nécessite aucune modification du flux binaire, un masquage neuronal des pertes de paquets et un schéma de redondance profonde (DRED) ; tous sont désactivés par défaut afin de ne pas alourdir les anciens appareils. Opus 1.6 a ajouté une extension neuronale de bande passante et un modèle DRED trois fois plus petit, tout en restant entièrement compatible avec la norme de 2012. L'approche est « dans une norme, optionnelle et rétrocompatible ».

Meta est allée plus loin dans l'autre sens. En juin 2024, elle a lancé MLow, un codec vocal à faible débit fondé sur le traitement du signal classique plutôt que sur un réseau, désormais utilisé dans tous les appels Instagram et Messenger et en cours de déploiement sur WhatsApp. Les raisons de Meta tenaient à ses appareils : plus de 20 % de ses appels sont passés sur des téléphones ARMv7, et des dizaines de millions d'appels WhatsApp par jour passent par des téléphones de plus de dix ans. Les codecs neuronaux, y compris EncCodec de Meta, ne fonctionnent de manière fiable que sur des téléphones coûteux, a écrit Meta. MLow offre environ deux fois la qualité d'Opus à 6 kbps tout en utilisant 10 % de calcul en moins.

Pour un service dont les utilisateurs vont des ordinateurs portables haut de gamme aux téléphones vieux de dix ans, c'est le budget de complexité qui décide. Les codecs neuronaux sont déployés d'abord là où le parc installé est récent et où le fournisseur contrôle les deux extrémités.

Ce que cela signifie pour les entreprises des communications

La visioconférence d'entreprise, la téléprésence, l'assistance à distance, les liaisons de contribution et les réseaux privés sont les endroits où un codec neuronal de bout en bout est aujourd'hui un produit, à condition que les deux extrémités vous appartiennent. Prévoyez trois choses dès le départ : un codec de secours, des tests de décodage sur chaque NPU que vous prenez en charge, et une évaluation de la qualité sur votre propre contenu plutôt que sur le jeu de données d'un fournisseur.

CHAPITRE 05

IA GÉNÉRATIVE

05 IA générative : le plus grand déploiement, que personne n'appelle un codec

Chaque modèle vocal et vidéo repose sur un codec appris

Le plus grand déploiement de compression neuronale de bout en bout ne se trouve dans aucune norme de codec. Il se trouve au cœur de l'IA générative. Un modèle de parole n'écoute pas de formes d'onde et n'en produit pas directement : il lit et écrit les jetons discrets d'un codec audio appris. Mimi, de Kyutai, publié en 2024, transforme la parole en jetons à 12,5 trames par seconde et 1,1 kbps, moins que n'importe quel codec de téléphonie, en combinant informations sémantiques et acoustiques dans un seul flux. EnCodec, de Meta, publié en 2022, en est un autre ; les propres ingénieurs de Meta ont ensuite décrit ces codecs comme excellents à bas débit mais trop lourds pour les téléphones plus anciens (chapitre 4).

La génération vidéo fonctionne de la même façon. Un modèle de diffusion ou un transformeur génère dans l'espace latent d'un autoencodeur vidéo appris (VAE), et le décodeur du VAE transforme le résultat en pixels. Le VAE de LTX-Video compresse la vidéo dans un rapport de 1:192 : chaque jeton latent représente un bloc de 32×32 pixels sur 8 images, avec 128 canaux. Son décodeur effectue aussi l'étape finale de débruitage, si bien qu'il génère en partie du détail au lieu de seulement le reconstruire.

Pourquoi ce ne sont pas des formats de distribution

Ces codecs s'exécutent à chaque requête adressée à un modèle vocal ou vidéo, mais aucun n'est un format que quelqu'un d'autre décode. Chaque modèle a son propre espace latent ; il n'y a ni régulation de débit, ni test de conformité, ni garantie qu'un autre décodeur produirait la même image. Ils sont conçus pour la qualité de généra-

tion, pas pour la fidélité à un débit donné, et ils décodent sur des GPU de centres de données. Leur succès montre que la compression apprise fonctionne à grande échelle quand une entreprise possède l'encodeur, le décodeur et le matériel.

Ce que cela signifie pour les entreprises

Deux points pratiques. D'abord, l'efficacité du tokeniseur est une ligne de coût directe pour quiconque fournit de la vidéo ou de la voix générative : pour un transformeur dont l'attention porte sur tous les jetons, le calcul croît plus vite que le nombre de jetons, si bien qu'un latent plus compact réduit le coût de génération plus que proportionnellement analyse. Les ingénieurs codec qui maîtrisent la conception des latents, les modèles entropiques et l'efficacité des décodeurs sont précieux pour les équipes d'IA générative. Ensuite, la frontière se déplace : les organismes de normalisation ont commencé à travailler sur un codage destiné aux machines plutôt qu'aux humains. Le Video Coding for Machines de MPEG (ISO/IEC 23888-2) a atteint le stade de projet de norme internationale début 2026 et sa version finale est prévue pour fin 2026, et JPEG AI a été conçu dès le départ pour servir les tâches de vision par ordinateur autant que le visionnage humain.

CHAPITRE 06

JEUX ET GRAPHISME

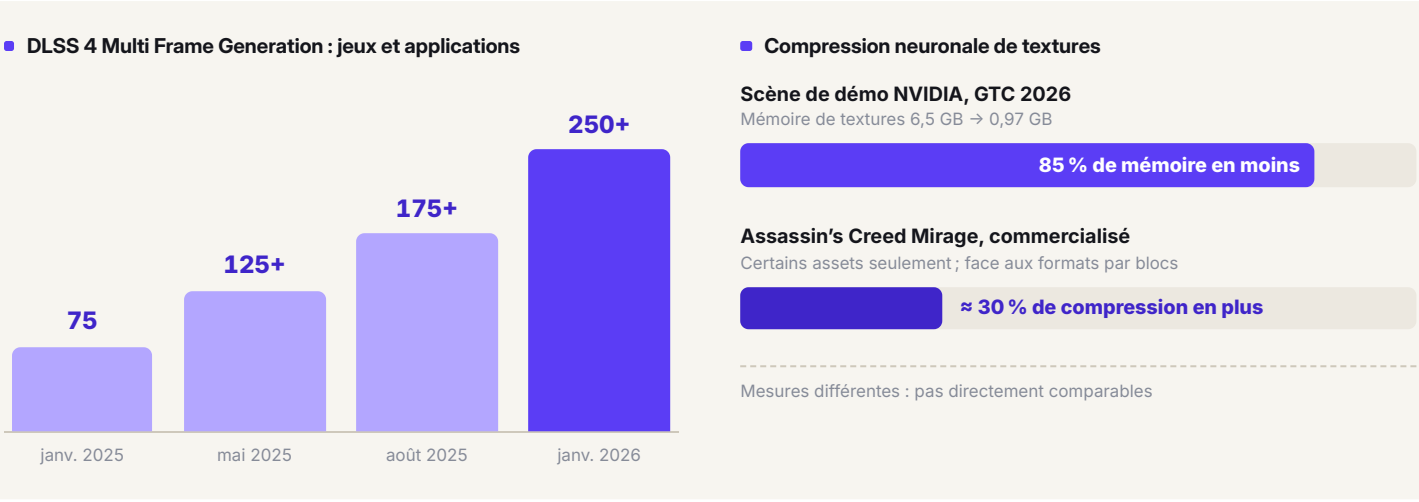
06 Jeux et graphisme : la reconstruction neuronale est généralisée, les assets neuronaux démarrent

Rendre moins de pixels et laisser un réseau compléter

Dans le graphisme temps réel, un réseau reconstruit déjà l'essentiel de ce que voient les joueurs. Le DLSS de NVIDIA effectue le rendu de moins de pixels et de moins d'images, et utilise des réseaux pour produire le reste. DLSS 4 avec Multi Frame Generation a été lancé en janvier 2025 avec une prise en charge dans 75 jeux et applications ; NVIDIA en comptait plus de 125 en mai 2025, plus de 175 en août et plus de 250 en janvier 2026, lorsque la société a annoncé DLSS 4.5 avec un modèle transformeur de deuxième génération pour la super-résolution. C'est un décodeur en tout sauf le nom : le « flux binaire » est une image basse résolution accompagnée de données de mouvement, et le réseau reconstruit l'image complète.

FIG. 05 **Quand une entreprise contrôle toute la pile, la reconstruction neuronale se diffuse vite**

À gauche : jeux et applications prenant en charge DLSS 4 Multi Frame Generation de NVIDIA. À droite : compression neuronale de textures dans une scène de démonstration et dans un jeu commercialisé, qui ne l'a utilisée que sur certains assets en raison de son coût à l'exécution.



Source: NVIDIA (janvier 2025 à janvier 2026 ; démo GTC 2026) ; Ubisoft La Forge (février 2026). Données des fournisseurs ; les deux chiffres sur les textures ne mesurent pas la même chose.

Les textures neuronales ont été livrées, de façon sélective

Les textures sont compressées dans des formats par blocs fixes (BCn) que les GPU décodent en matériel. La compression neuronale de textures remplace un ensemble de textures de matériau par une représentation latente compacte et un petit réseau qui reconstruit les texels au moment où ils sont nécessaires. En février 2026, le laboratoire La Forge d'Ubisoft a décrit son intégration dans Assassin's Creed Mirage, un jeu de 2023, avec une compression effective environ 30 % supérieure à celle des formats conventionnels à qualité visuelle égale selon le fournisseur. Elle n'a pas été utilisée partout : « l'usage était limité par les performances à l'exécution », si bien qu'Ubisoft l'a appliquée à un sous-ensemble d'assets aux nombreuses instances et soumis à une forte pression mémoire.

Le RTX Neural Texture Compression SDK de NVIDIA, en bêta sur GitHub, montre le potentiel maximal. Lors d'une démonstration au GTC 2026, une scène de villa qui utilisait 6,5 GB de mémoire de textures avec BCn n'en nécessitait plus que 970 MB avec des textures neuronales, soit 85 % de moins selon le fournisseur. Le SDK propose deux voies : décoder en BCn au chargement, ce qui économise de l'espace disque et du volume de téléchargement mais pas de mémoire vidéo, ou décoder à l'échantillonnage, ce qui économise la mémoire mais coûte du temps de shader à chaque image.

L'API arrive

Les API graphiques ajoutent les opérations matricielles dont le décodage neuronal a besoin dans les shaders. Microsoft a livré Shader Model 6.9 en février 2026 ; la fonctionnalité de vecteurs coopératifs présentée en préversion en 2025 a été intégrée à une conception unifiée d'algèbre linéaire destinée à un shader model ultérieur, DirectX Linear Algebra entrant en préversion publique en avril 2026. Lorsque cela figurera dans tous les pilotes, un décodeur de textures neuronales pourra fonctionner sur le GPU de n'importe quel fournisseur à partir d'un seul chemin de code.

Ce que cela signifie pour les entreprises

Les jeux déploient le décodage neuronal parce que les conditions d'une boucle fermée sont réunies : un seul moteur, une seule chaîne de production d'assets, du matériel tensoriel dans le GPU et un format de secours. Les

chantiers ouverts portent sur les outils, les chaînes de production d'assets qui encodent des matériaux neuronaux, les environnements d'exécution multifournisseurs et, surtout, le mobile, où le coût énergétique par texel décodé est décisif.

CHAPITRE 07

07

LIAISONS SPÉCIALISÉES

Liaisons spécialisées : satellites, drones et robots

Ce sont les cas où le codage neuronal de bout en bout convient le mieux sur le papier. Un seul propriétaire contrôle les deux extrémités de la liaison, la bande passante est rare et sujette aux pertes, le contenu est assez restreint pour se prêter à l'entraînement d'un modèle, et une image fidèle sur le plan perceptif, voire seulement sur le plan sémantique, suffit souvent. Les codecs neuronaux résistants aux pertes pour ce type de liaisons sont un domaine de recherche actif, et la restauration côté serveur de flux à faible débit est utilisée.

Ils constituent aussi un marché petit, et de plus en plus réglementé, comparé à la vidéo grand public analyse. En Chine, l'annonce n° 49 de 2026 du ministère du Commerce et de deux autres organismes, publiée le 19 octobre et en vigueur le 18 novembre 2026, révise le contrôle des exportations de 13 catégories d'articles liés aux drones. Parmi les rubriques technologiques, l'une couvre les plateformes terrestres sans pilote, définies en partie par des seuils tels qu'une compression sémantique d'image de 75 % et une reconstruction d'image en 150 millisecondes au plus, selon la presse spécialisée. La compression sémantique est elle-même devenue un paramètre du contrôle des exportations. Aux États-Unis, la Federal Communications Commission a ajouté le 22 décembre 2025 les drones de fabrication étrangère et leurs composants critiques à sa Covered List, si bien que les nouveaux modèles étrangers ne peuvent plus obtenir d'autorisation d'équipement.

Le verdict : un créneau solide pour une entreprise spécialisée ayant des clients industriels ou publics et la capacité de gérer la conformité à l'export des deux côtés, mais pas un marché de masse, ni l'endroit où apprendre si le codage neuronal fonctionne à grande échelle.

CHAPITRE 08

SILICIUM

08 Le point de vue du silicium : énergie, calcul, mémoire et exactitude

C'est le chapitre où mon propre parcours compte le plus. J'ai derrière moi de nombreuses années de R&D et de management dans les puces fabless, dont la conception full-custom de chemins de données d'ASIC et des projets de puces multimédias pour caméras, comme des accélérateurs de codec vidéo, et les questions qu'une équipe puce se pose sur un nouveau codec ne portent pas sur le BD-rate. Les voici : combien de joules par pixel, combien d'octets de trafic DRAM par pixel, combien de portes, et chaque décodeur au monde peut-il calculer exactement les mêmes bits ? Sur ces quatre points, un décodeur neuronal en est en 2026 là où en était le HEVC matériel avant ses premières puces : prometteur en logiciel, coûteux en silicium.

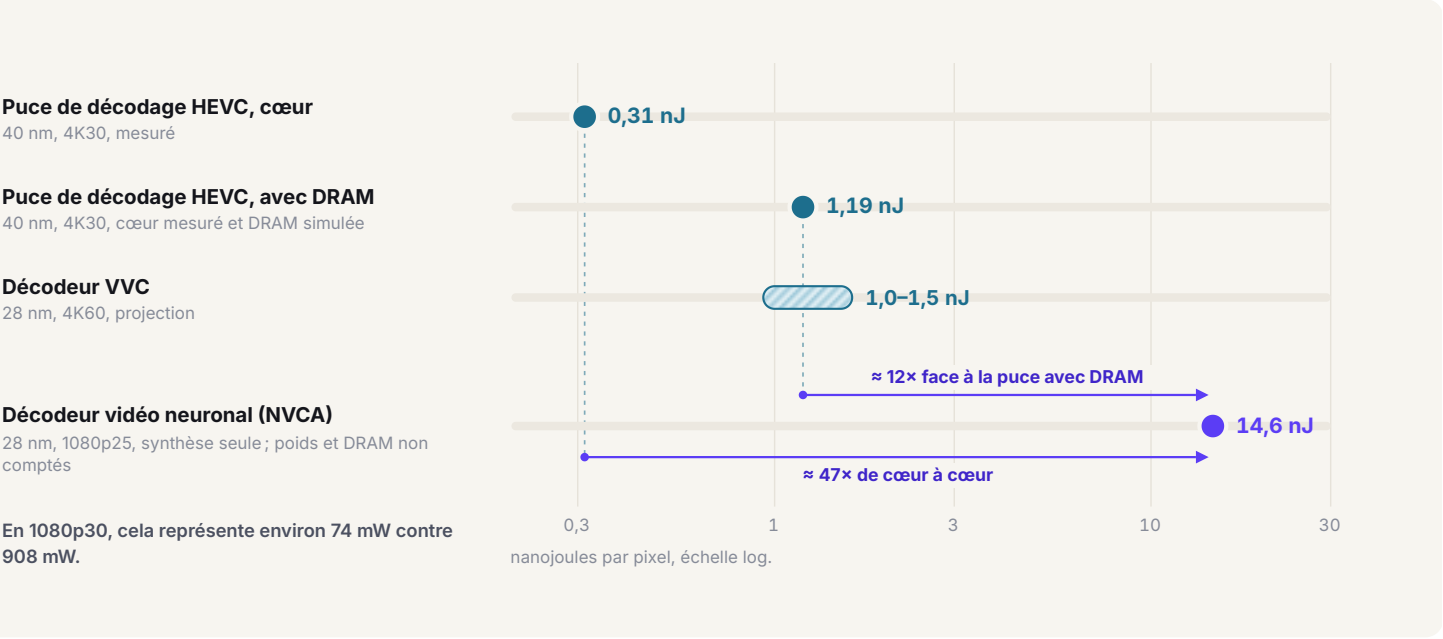
L'énergie par pixel : un ordre de grandeur d'écart

Peu de décodeurs neuronaux ont été conçus sous forme de puces, si bien que la comparaison repose sur une poignée de conceptions publiées. La référence côté classique est la puce de décodage HEVC du MIT (Tikekar et al., ISSCC 2013 et IEEE JSSC 2014) : 40 nm, 1,77 mm², décodage de la 4K à 30 images par seconde à 200 MHz. Son cœur consomme 78 mW, soit 0,31 nJ par pixel ; en comptant la consommation simulée de la DRAM, le système nécessite 1,19 nJ par pixel. Pour VVC, les chiffres de production sont confidentiels ; une analyse parue en 2026 dans Electronic Design, qui transpose la puce du MIT en 28 nm et applique la complexité plus élevée de VVC, projette 1–1,5 nJ par pixel pour un décodeur 4K60 d'environ un million de portes.

Côté neuronal, la seule conception d'ASIC de décodeur vidéo neuronal de bout en bout publiée, selon le décompte d'Electronic Design, est NVCA (Zhang et al., DATE 2024), synthétisée en TSMC 28 nm avec 5,01 millions de portes, sans compter ses poids, stockés hors puce. Elle décode du 1080p à 25 images par seconde pour 760 mW, soit 14,6 nJ par pixel.

FIG. 06 **Énergie par pixel décodé : les décodeurs à fonction fixe ont un ordre de grandeur d'avance**

La seule conception publiée de puce de décodage vidéo neuronal exige environ 12 fois l'énergie par pixel d'une puce de décodage HEVC mesurée, DRAM comprise, et environ 47 fois celle de son seul cœur. Le trafic des poids et de la DRAM propre à la conception neuronale n'est pas compté.



Source: Tikekar et al., IEEE JSSC 2014 (MIT) ; NVCA, Zhang et al., DATE 2024 (arXiv 2312.10716) ; la fourchette VVC est une projection d'Electronic Design (2026). La puce HEVC utilise un nœud plus ancien (40 nm contre 28 nm), ce qui avantage la conception neuronale.

C'est environ 12 fois la puce HEVC comptée avec sa DRAM, et environ 47 fois son seul cœur. La comparaison avantage NVCA de trois façons : la puce HEVC utilise un nœud plus ancien, le chiffre de NVCA laisse de côté le trafic de ses poids et de la DRAM, et NVCA est un résultat de synthèse et non du silicium mesuré. En 1080p30, l'écart est d'environ 74 mW contre environ 900 mW : un décodeur que personne ne remarque, face à un décodeur qui pèse sur la batterie et le budget thermique d'un téléphone.

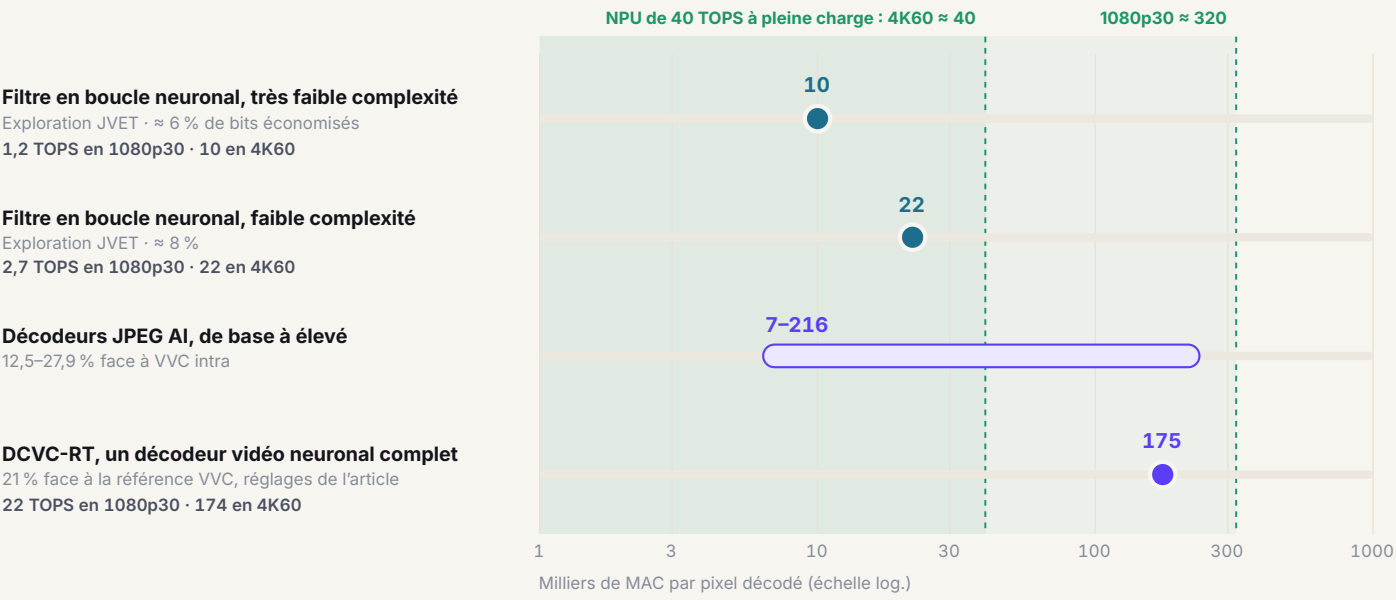
Le calcul par pixel : ce qu'un décodeur neuronal demande à un NPU

Le codec vidéo neuronal publié le plus rapide que j'aie trouvé est DCVC-RT de Microsoft Research (Jia et al., CVPR 2025). Il encode et décode du 1080p à 125 et 113 images par seconde sur un GPU NVIDIA A100 tout en économisant 21 % de débit par rapport à l'encodeur de référence de VVC dans les conditions de test de l'article, et il utilise l'arithmétique entière pour que différents appareils décodent de façon cohérente. Ses documents complémentaires indiquent la complexité de chaque module. En additionnant les quatre modules qu'un décodeur doit exécuter (décodeur, extracteur de caractéristiques, modèle entropique et reconstruction), j'obtiens environ 175 000 multiplications-accumulations par pixel décodé estimation.

En 1080p30, cela représente environ 11 000 milliards de multiplications-accumulations par seconde, soit environ 22 TOPS à raison de deux opérations par MAC, avant toute perte d'utilisation ; la 4K60 en demande huit fois plus estimation. Pour donner l'échelle, Microsoft définit ses PC IA par un NPU d'au moins 40 TOPS, et Apple annonce 38 TOPS pour le Neural Engine du M4. Ce sont des chiffres de crête, partagés avec tout ce que fait par ailleurs l'appareil.

FIG. 07 **Un outil neuronal coûte des dizaines de milliers de MAC par pixel ; un décodeur neuronal complet, des centaines de milliers**

Multiplications-accumulations du décodeur par pixel, en échelle logarithmique, comparées à ce qu'un NPU de 40 TOPS pleinement utilisé pourrait absorber en 4K60 et en 1080p30. Les gains sont mesurés par rapport à des références différentes et ne sont pas directement comparables.



Source: JVET, via le président du SC 29 (avril 2026) ; comité JPEG (2024) ; DCVC-RT, documents complémentaires CVPR 2025, somme des modules du décodeur par l'auteur ; définition des PC Copilot+ de Microsoft. Calcul de l'auteur : deux opérations par MAC, pleine utilisation.

Les échelons de la figure 7 offrent la bonne grille de lecture du domaine. Les outils neuronaux qui coûtent des dizaines de milliers de MAC par pixel, comme les filtres en boucle du JVET (environ 6 % d'économie pour 10 000, 8 % pour 22 000), sont plausibles dans un futur décodeur normalisé. JPEG AI couvre toute la plage, de 7 000 à 216 000 MAC par pixel pour 12,5–27,9 % de mieux que VVC intra. Un décodeur vidéo neuronal complet, à quelque deux cent mille MAC par pixel, a sa place sur un NPU, à résolution modeste, en boucle fermée. C'est exactement là que se situe MLVC : 540p30, en utilisant au plus la moitié du NPU.

Le trafic mémoire, plus que les MAC, est souvent la vraie limite

Les auteurs de DCVC-RT ont constaté que ce qui limitait la vitesse n'était pas l'arithmétique mais le « coût opérationnel » : les entrées-sorties mémoire et le nombre d'appels de fonction. C'est en grande partie pour cette raison qu'ils ont supprimé les modules de mouvement explicites et le traitement multi-échelle. Les concepteurs de NVCA ont construit un flux de données spécifique pour réduire le trafic mémoire hors puce lié aux caractéristiques de mouvement et de résidu. Un concepteur de puces reconnaîtra le schéma. Dans la puce HEVC du MIT, la DRAM représente environ les trois quarts de l'énergie du système (0,88 sur 1,19 nJ par pixel), même après un cache qui a réduit la bande passante DRAM de 67 % analyse. Un décodeur neuronal qui conserve de grandes cartes de caractéristiques multicanaux d'une image à l'autre aggrave le problème de mémoire au lieu de l'atténuer.

L'exactitude : chaque décodeur doit calculer les mêmes bits

Le codage entropique ne fonctionne que si le modèle de probabilité du décodeur correspond exactement à celui de l'encodeur. Dans un codec classique, ce modèle est une table ou un petit ensemble de règles entières. Dans un codec neuronal, c'est un réseau, et une différence de calcul en virgule flottante entre deux appareils, due à un

autre noyau de calcul, à un autre ordre d'accumulation ou à un autre NPU, peut amener le décodeur à lire de mauvais symboles et à perdre le flux à partir de ce point.

Tous ceux qui ont livré ou normalisé un codec neuronal ont dû éliminer ce problème par l'ingénierie. DCVC-RT convertit son modèle en entiers pour un codage cohérent d'un appareil à l'autre. MLVC utilise une conception à hyperprior à échelle partagée pour résister aux écarts entre plateformes, ce qui explique probablement en partie pourquoi il cède de la compression par rapport à DCVC-RT analyse. L'exploration des réseaux de neurones du JVET fait passer chaque réseau par SADL, une petite bibliothèque d'inférence en entiers 16 bits. JPEG AI définit des points de conformité dans les domaines latent, résiduel et pixel. Pour une équipe puce, c'est la même discipline que pour les transformées entières de HEVC : la norme doit définir l'arithmétique exactement, ce qui contraint la conception du réseau et exclut les raccourcis propres à chaque fournisseur.

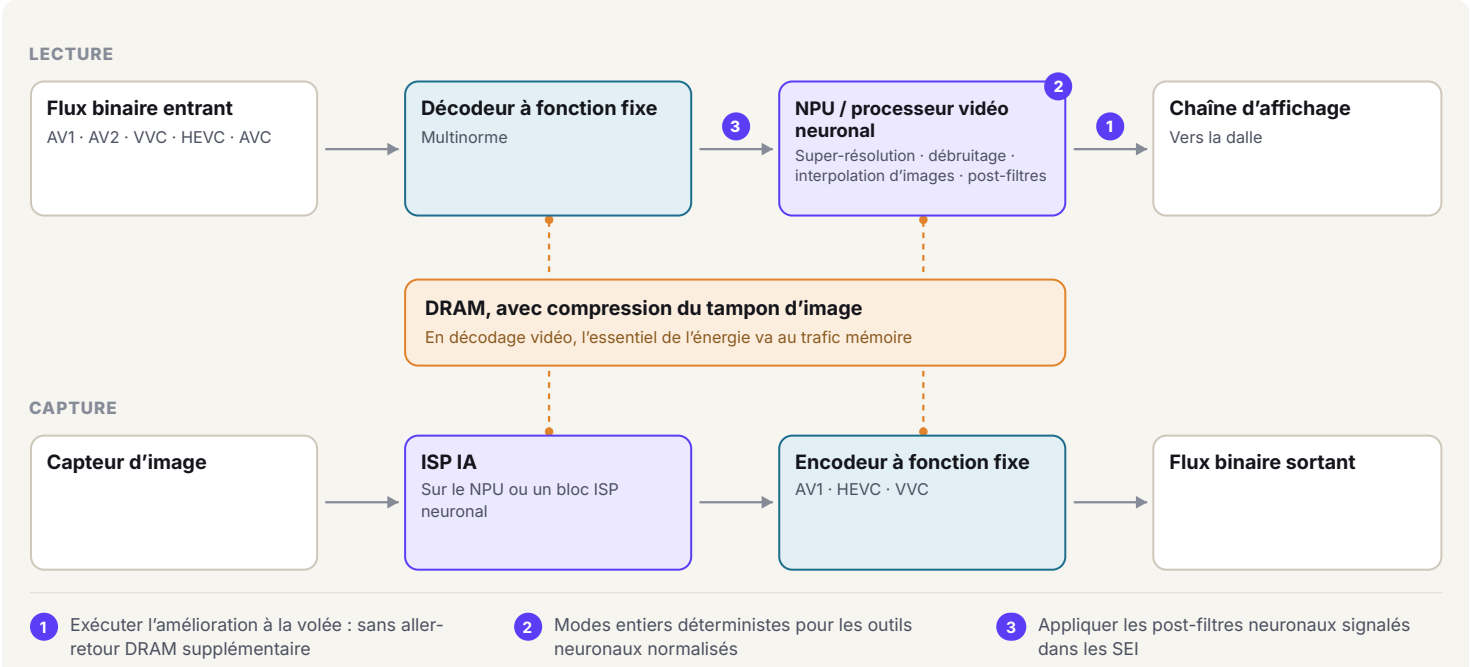
La surface n'est pas l'obstacle ; la vérification et la durée de vie le sont

Les 5 millions de portes de NVCA, contre environ un million pour un décodeur VVC projeté, paraissent beaucoup, mais sur un SoC de téléphone ou de téléviseur moderne qui embarque déjà un NPU de plusieurs TOPS, quelques millions de portes ne sont pas ce qui décide de la conception. Ce qui coûte vraiment cher, c'est la vérification face à une suite de tests de conformité, le maintien de l'exactitude d'un décodeur programmable au fil des mises à jour logicielles pendant toute la durée de vie d'un appareil, et la consommation en 4K60 HDR. Ces coûts n'ont de sens qu'une fois qu'une norme fixe ce qui doit être construit.

L'architecture qui l'emporte aujourd'hui : un codec à fonction fixe plus un NPU

FIG. 08 **Le partage des tâches dans les produits : du matériel à fonction fixe pour le flux binaire, un moteur neuronal pour les pixels**

Un SoC multimédia de 2026. Le bloc codec décode toutes les normes pour une fraction de nanojoule par pixel ; le NPU fait ce que les réseaux font bien et que les normes n'encadrent pas. Le trafic DRAM est le coût qui compte.



Source: Vue d'architecture de l'auteur, fondée sur les informations produits publiées par Tencent (Canghai V2), Chips&Media, Allegro DVT et Samsung (2026).

Le silicium qui se vend en 2026 associe un codec multinorme à fonction fixe à un moteur neuronal placé à côté. Le bloc codec traite AV1, AV2, VVC, HEVC et AVC pour une fraction de nanojoule par pixel. Le moteur neuronal traite ce que les réseaux font bien et que les normes n'encadrent pas : le débruitage et l'ISP à la capture, la super-résolution, l'interpolation d'images et l'amélioration à l'affichage. Les deux partagent la compression du tampon d'image et la planification de la DRAM, parce que le trafic mémoire est le coût qui compte.

Le marché montre ce partage. L'ASIC de codec vidéo Canghai (沧海) de Tencent, développé en interne depuis 2019, compte plus de 100 000 puces de première génération déployées dans Tencent Cloud et dans ses propres services de direct, de vidéo à la demande et de cloud gaming ; la deuxième génération, annoncée en mai 2026 comme mise sous tension et entrant en production en volume pour une mise en service au second semestre, prend en charge H.265 et H.266 dans une même architecture, avec une compression supérieure de plus de 10 % et de 30 % à celle de son prédécesseur, environ deux fois plus de débit de traitement, et une amélioration d'image sur puce comme la super-résolution. Chips&Media vend côte à côte des IP de codec et un NPU orienté vidéo, le WAVE-N. Allegro DVT vend des décodeurs AV2 et a ajouté sa première IP de traitement vidéo neuronal. Les processeurs TV de Samsung associent des décodeurs normalisés à des centaines de réseaux de mise à l'échelle.

Ce que je construirais, selon le rôle

Équipes SoC et puces TV. Conservez des décodeurs à fonction fixe pour chaque norme que vous devez lire, y compris l'AV2 dès maintenant et le VVC là où la télédiffusion l'exige. Dimensionnez le NPU pour le post-traitement à la résolution et à la fréquence d'images de l'affichage, et concevez le chemin du décodeur au NPU puis à l'affichage de sorte que l'amélioration n'ajoute pas d'aller-retour par la DRAM. Prenez en charge les SEI de post-filtre neuronal. Prévoyez des modes entiers déterministes dans le NPU. Considérez H.267 comme un point de départ de conception en 2029 ou après, et gardez une part de programmabilité pour ses outils neuronaux.

Fournisseurs d'IP. Vendez dès maintenant des décodeurs AV2 et VVC, plus les briques dont le traitement neuronal a besoin autour d'eux : accélérateurs de post-filtres, compression du tampon d'image et outils de conformité.

Plateformes vidéo. Le prétraitement neuronal à l'encodeur est rentabilisé en premier, parce qu'il s'exécute une fois par titre et n'exige rien des appareils. Utilisez la super-résolution côté client là où les appareils le permettent. Réservez les codecs neuronaux de bout en bout aux boucles fermées que vous contrôlez.

CHAPITRE 09

NORMES ET BREVETS

09 Normes et brevets : là où les outils neuronaux passent sous licence

Les normes en préparation

Norme	Organisme	Contenu neuronal	Statut, novembre 2026
H.267 (nom de travail ; au-delà de VVC)	JVET (ITU-T et ISO/IEC)	Ouvert. Outils classiques (ECM) et outils neuronaux (filtres en boucle, prédiction intra, super-résolution) sont tous candidats	Appel à propositions en juillet 2026 ; réponses évaluées en janvier 2027 ; premier modèle de test au second semestre 2027 ; version 1 fin 2029
VVC et HEVC, avec VSEI	JVET	Un post-filtre neuronal peut être décrit dans les métadonnées SEI ; son application est optionnelle pour le décodeur	En vigueur ; messages étendus dans les éditions d'octobre 2025
AV2	AOMedia	Outils classiques, sans redevances	Version 1.0.0 le 28 mai 2026 ; IP de décodage sous licence depuis juin 2026 ; flux Web YouTube prévus début 2027
JPEG AI	JPEG (ISO/IEC et ITU-T)	Codec d'image appris de bout en bout, pour la vision humaine et la vision par machine	Partie 1 publiée en 2025 ; une deuxième version en discussion ; aucune prise en charge par défaut trouvée dans les produits
IEEE 1857.11	IEEE	Codage d'image par réseaux de neurones conçu pour les NPU	Publiée en décembre 2024
Video Coding for Machines	MPEG	Codage pour l'analyse par machine plutôt que pour le visionnage	Projet de norme internationale début 2026 ; version finale prévue fin 2026
Extensions ML d'Opus	IETF	Redondance profonde, amélioration côté décodeur et extension de bande passante, toutes optionnelles	Dans libopus 1.5 et 1.6 ; redondance profonde en cours de normalisation

Deux choses ressortent. D'abord, le contenu neuronal entre dans les normes par les marges avant d'atteindre le cœur : sous forme de post-filtres optionnels, d'amélioration optionnelle et de codage d'image, qui laissent tous inchangé le chemin de décodage principal. Ensuite, la seule norme qui pourrait placer des outils neuronaux dans le chemin de décodage de tous les écrans, H.267, est à trois ans d'une première version.

Qui est positionné sur les brevets

Les outils d'une norme deviennent des brevets sous licence ; le processus H.267 est donc aussi une course aux brevets, et les outils neuronaux en font partie. La cartographie publiée par GreyB en 2025, qui porte sur 3 916 brevets post-VVC, place ByteDance en tête, avec 645 familles de brevets et 251 contributions au JVET, et relève des partenariats université–industrie comme celui de Xiaomi et de l'université Xidian, avec plus de 20 familles. MLex a rapporté en mars 2026 que Tencent, Alibaba et ByteDance s'affirment comme des participants actifs des pools de brevets de codecs vidéo. En mai 2026, Sauvegarder Investment Management a acheté à Alibaba 126 ac-

tifs de brevets de codecs vidéo couvrant AV1, AV2 et Video Coding for Machines. Le rachat de Deep Render par InterDigital en octobre 2025 a ajouté à son portefeuille les brevets de codage vidéo par IA de Deep Render.

L'argent par appareil se lit facilement dans les grilles tarifaires des pools. Les tarifs d'Access Advance pour VVC (au 1er juillet 2026, pour les licenciés en conformité qui utilisent sa marque) sont de 0,50 dollar par téléphone mobile, tablette ou ordinateur portable sur les marchés de la Région 1 comme les États-Unis, l'UE et le Japon, et de 0,25 dollar en Région 2 ; pour les téléviseurs 4K, de 1,00 à 1,50 dollar selon le prix en Région 1 ; avec un plafond annuel par entreprise de 75 millions de dollars. AV1 et AV2 sont exempts de redevances de la part des membres d'AOMedia. Quel que soit le contenu de H.267, ses outils seront tarifés de la même façon, et c'est pourquoi les entreprises qui détiennent des brevets de codage neuronal contribuent dès maintenant au JVET analyse.

CHAPITRE 10

QUI GAGNE DE L'ARGENT

10 Qui gagne de l'argent, et où une entreprise peut se positionner

L'argent en 2026

FIG. 09 **Qui gagne de l'argent avec le codage neuronal en 2026**

L'argent suit les normes et les plateformes. Je n'ai trouvé aucune entreprise indépendante qui vive de la concession de licences d'un codec vidéo neuronal de bout en bout.

REVENUS ACTUELS

Plateformes

Gardent les économies

Sous-échantillonneur neuronal de Netflix ; super-résolution chez YouTube et Meta ; MLVC publié en open source par Microsoft sous licence MIT

REVENUS ACTUELS

Fabricants d'appareils et de puces

Vendent de la différenciation

ISP des appareils photo ; processeurs de mise à l'échelle pour téléviseurs ; puces codec Canghai maison de Tencent

REVENUS ACTUELS

Fournisseurs d'IP

Vendent des licences à chaque nouvelle norme

Clients pour des décodeurs AV2 chez Chips&Media, Allegro DVT et VeriSilicon dans les trois semaines suivant la spécification

REVENUS ACTUELS

Fournisseurs de GPU

Créent une dépendance à leur plateforme

DLSS 4 dans plus de 250 jeux et applications ; SDK de textures neuronales en bêta

PAR APPAREIL

Détenteurs de brevets et pools

Perçoivent par appareil

VVC : 0,50 $ par téléphone, 1,00–1,50 $ par TV 4K ; outils H.267 après 2029

SORTIE PAR RACHAT

Start-up de codecs neuronaux

Rachetées

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

Source: Publications d'entreprises et de pools de brevets citées dans le rapport ; situation au 12 novembre 2026. Tarifs du pool : Access Advance, VVC, Région 1, licencié en conformité utilisant la marque, au 1er juillet 2026.

Les plateformes gardent les économies. Netflix, YouTube, Meta, Microsoft et Tencent développent le traitement neuronal en interne et en conservent le bénéfice sous forme de bande passante réduite, de stockage réduit et d'une meilleure qualité par bit. Quand cela favorise l'adoption, elles donnent le code : Microsoft a publié MLVC sous licence MIT.

Les fabricants d'appareils et de puces vendent de la différenciation. Les ISP des appareils photo et les processeurs d'image des téléviseurs sont l'endroit où les réseaux de neurones touchent chaque utilisateur, et ils sont vendus comme des fonctionnalités : le Cognitive ISP de Qualcomm, les processeurs de mise à l'échelle de Samsung.

Les fournisseurs d'IP vendent des décodeurs pour chaque nouvelle norme, plus des NPU. Dans les trois semaines qui ont suivi la spécification de l'AV2, Chips&Media, Allegro DVT et VeriSilicon ont tous annoncé des clients pour leurs décodeurs AV2. Les mêmes fournisseurs vendent désormais des blocs de traitement neuronal à côté des codecs.

Les fournisseurs de GPU construisent des plateformes. Le DLSS et la compression neuronale de textures font du matériel tensoriel de NVIDIA un élément de la fabrication même des jeux.

Les détenteurs de brevets perçoivent des redevances par appareil, via des pools pour VVC et HEVC et, après 2029, pour H.267.

Les start-up de codecs neuronaux ont été rachetées. WaveOne a rejoint Apple en 2023, iSIZE Sony Interactive Entertainment en 2023, et Deep Render InterDigital en 2025, toutes à des conditions non divulguées. Dans chaque cas, une plateforme ou un concédant de licences a acheté l'équipe et les brevets. Je n'ai trouvé aucune entreprise indépendante qui, en 2026, vive de la concession de licences d'un codec vidéo neuronal de bout en bout.

Où une entreprise peut se positionner

Créneau	Qui achète	Pourquoi maintenant	Risque principal
Prétraitement neuronal à l'encodeur (sous-échantillonnage, débruitage, réglage par titre)	Plateformes de streaming et de vidéos d'utilisateurs, diffuseurs	Aucun changement d'appareil nécessaire ; éprouvé chez Netflix	Les grandes plateformes le développent elles-mêmes
IP et réglage d'amélioration côté affichage (super-résolution, interpolation d'images, qualité d'image)	Fabricants de puces pour téléviseurs, boîtiers TV et téléphones	Les NPU sont déjà dans les appareils ; différenciation visible	Banalisation ; développement interne chez les fabricants de puces
Codecs neuronaux en boucle fermée (visioconférence, opérations à distance, liaisons privées)	Fournisseurs de communications d'entreprise, industriels et de téléprésence	MLVC prouve la faisabilité, et son code est ouvert	Il faut contrôler les deux extrémités ; fragmentation des NPU
Silicium ou IP codec plus NPU (AV2, VVC et post-traitement)	Fabricants de SoC	Le cycle AV2 commence	Acteurs établis puissants
Outils d'évaluation, de conformité et d'inférence déterministe	Quiconque livre des outils neuronaux	L'exactitude au bit près est une exigence stricte	Petit marché, sauf en offre groupée
Contributions normatives à H.267	Concédants de licences, grandes plateformes	Les réponses seront évaluées en janvier 2027	Long, coûteux, incertain

Pour la plupart des entreprises, les deux premières lignes sont celles où des revenus sont accessibles en 2027. La troisième est la plus intéressante pour une entreprise produit ciblée. La dernière exige de la patience et une stratégie de licences.

CHAPITRE 11

AFFIRMATIONS ET FAITS

11 Les affirmations que vous entendrez, et ce que disent les faits

Affirmation	Ce que disent les faits
« Les codecs neuronaux battent déjà VVC. »	En laboratoire, oui. DCVC-RT économise 21 % par rapport à l'encodeur de référence de VVC en 1080p dans les conditions de son article ; JPEG AI économise 12,5–27,9 % par rapport à VVC intra. Le décodeur coûte environ 175 000 MAC par pixel, soit l'énergie d'un NPU, contre une fraction de nanojoule en matériel à fonction fixe.
« Les codecs IA remplaceront H.26x et AV1 d'ici quelques années. »	Pas dans la lecture vidéo. La première norme capable d'embarquer des outils neuronaux vise fin 2029, les puces suivront vers 2031 et l'usage se généralisera au milieu des années 2030 estimation. Le codec qui arrive maintenant, l'AV2, est classique.
« MLVC réduit la bande passante de 88 %. »	Par rapport à H.264 en 360p, sur le propre jeu de test de Microsoft. Par rapport au HEVC matériel, c'est 75,5 % en 360p et 65,4 % en 540p, dans des appels où Microsoft contrôle les deux extrémités.
« Les textures neuronales réduisent la mémoire vidéo de 85–96 %. »	Dans des scènes de démonstration. Le jeu commercialisé a gagné environ 30 % de compression effective et n'a utilisé les textures neuronales que sur certains assets, en raison de leur coût à l'exécution.
« JPEG AI est une norme ISO, donc les appareils photo et les navigateurs vont l'adopter. »	Je n'ai trouvé aucune prise en charge par défaut dans les produits. Le format adopté par le Web en 2026, JPEG XL dans Chrome 155, est classique.
« Le décodage neuronal est trop coûteux pour les téléphones. »	En partie vrai. MLVC décode du 540p30 avec la moitié d'un NPU ; la 4K60 avec un modèle de classe DCVC-RT exigerait environ 175 TOPS de calcul soutenu estimation.
« Les liaisons satellite et drone sont le grand marché de la vidéo neuronale. »	Techniquement le meilleur terrain, mais un marché petit et de plus en plus soumis au contrôle des exportations des deux côtés du Pacifique.
« Les grandes plateformes prendront un codec neuronal sous licence auprès d'une start-up. »	Elles ont construit le leur (Netflix, Microsoft, Tencent) ou racheté la start-up (Apple, Sony, InterDigital).

Un test simple permet de trancher la plupart des argumentaires : demandez la référence, la métrique, la résolution et le coût de décodage. Une affirmation qui ne peut pas préciser ces quatre éléments est un résultat de recherche, pas un produit.

CHAPITRE 12

12 JALONS À surveiller, de 2027 à 2030

Quand	Jalon	Pourquoi c'est important
Janvier 2027	Le JVET évalue les réponses H.267 à Brisbane	Combien des meilleures réponses reposent sur des outils neuronaux, et à quel coût par pixel
Début 2027	Premiers flux Web AV2 de YouTube ; décodeurs AV2 dans les SoC annoncés	Le prochain cycle classique fixe les budgets des appareils jusque vers 2030
Second semestre 2027	Premier modèle de test de H.267	Si les outils neuronaux figurent dans la conception de base ou seulement dans l'exploration
2027	MLVC au-delà des appels Teams pair à pair ; d'autres fournisseurs adoptent le code ouvert	Le premier test d'un codec neuronal ouvert en dehors de son propriétaire
2027–2028	Un premier produit qui écrit ou lit JPEG AI par défaut	Le premier format neuronal utilisé au-delà des frontières d'une entreprise
2027–2028	DirectX Linear Algebra dans les pilotes publiés ; textures neuronales dans davantage de jeux et sur mobile	Le décodage neuronal multifournisseur au sein du graphisme
En continu	Budgets NPU des téléphones et des PC	À environ 100 TOPS soutenus, un décodeur 1080p30 de classe DCVC-RT occuperait environ un cinquième du NPU au lieu de la moitié analyse
Fin 2029	Version 1 de H.267	Le début du cycle silicium pour les outils neuronaux dans la lecture grand public

CHAPITRE 13

13 Questions ouvertes

QUESTIONS OUVERTES

Plusieurs questions abordées dans ce rapport n'ont pas encore de réponse publique, et ce sont celles que je trouve techniquement les plus intéressantes.

L'exactitude à grande échelle. Jusqu'où un décodeur neuronal peut-il s'approcher d'une arithmétique exacte au bit près, uniquement entière, sur des NPU de différents fournisseurs avant de perdre l'essentiel de son avantage en compression ? Le compromis de six points de MLVC est un élément de réponse ; une norme a besoin d'une réponse générale.

La mémoire avant les MAC. Que coûte le post-traitement neuronal en trafic DRAM par pixel lorsqu'il s'exécute à la volée entre un décodeur matériel et l'affichage, et jusqu'où peut-on réduire son ensemble de travail ?

Le prix d'un outil neuronal dans H.267. À partir de combien de MAC et de nanojoules par pixel un filtre en boucle neuronal ne vaut-il plus les bits qu'il fait économiser dans un décodeur 4K60 ?

Les boucles fermées au-delà des appels. Quels autres systèmes fermés, du stockage de photos dans le cloud et du cloud gaming aux robots et aux liaisons de contribution privées, peuvent adopter des codecs neuronaux de bout en bout avant l'arrivée de H.267 ?

Les décodeurs qui génèrent. Lorsqu'un décodeur génère du détail au lieu de le reconstruire, comme le font déjà les VAE vidéo, comment mesurer et spécifier la fidélité ?

Ce sont des domaines que nous pouvons explorer ensemble dans les années à venir, par des échanges techniques et une collaboration technique : i@tanshuai.com.

Limites de ce rapport

Ce rapport est une recherche documentaire fondée sur des documents publics disponibles jusqu'au 12 novembre 2026. Je n'ai mené ni entretien primaire ni mesure indépendante pour cette édition. Plusieurs chiffres clés sont des affirmations de fournisseurs ou d'auteurs sur leurs propres jeux de test, comme les économies de débit de MLVC et les taux de compression des textures neuronales ; ils sont signalés comme tels. Mes estimations de calcul pour DCVC-RT additionnent les valeurs par module publiées par ses auteurs et ignorent les pertes d'utilisation, si bien que la charge réelle du NPU sera plus élevée. La comparaison énergétique repose sur quelques conceptions publiées, sur des nœuds de gravure différents ; les chiffres des décodeurs VVC et AV2 de production sont confidentiels. Les calendriers de normalisation sont les objectifs des comités eux-mêmes. Les dates de H.267 postérieures à 2029 sont mes estimations, fondées sur les historiques de l'AV1 et du VVC.

À propos de l'auteur

Tan Shuai est un technologue et fondateur qui compte plus de vingt ans de travail dans les puces, le matériel et le logiciel système, et 20 brevets délivrés. Il a derrière lui de nombreuses années de R&D et de management dans les puces fabless, dont la conception full-custom de chemins de données d'ASIC et des projets de puces multimédias pour caméras, comme des accélérateurs de codec vidéo, ce qui constitue la grille de lecture de ce rapport : énergie par pixel, trafic mémoire, et ce qu'une norme permet à un décodeur de figer dans le silicium. Son parcours l'a mené dans la Silicon Valley, à New York, à Chicago, à Taïwan, dans la région de la Grande Baie (Shenzhen et Hong Kong), au Vietnam et en Malaisie. Il travaille aujourd'hui comme chercheur et développeur en IA.

Échanges techniques : i@tanshuai.com · tanshuai.com

Notes et avertissement

Ce rapport est une recherche documentaire indépendante fondée sur des documents publics disponibles jusqu'au 12 novembre 2026. Il est fourni à titre d'information uniquement et ne constitue pas un conseil en investissement, un conseil juridique ou un conseil d'ingénierie. Les produits, les calendriers de normalisation, les tarifs des pools de brevets et les règles d'exportation évoluent ; vérifiez auprès de la source primaire tout élément sur lequel vous vous appuyez. Les chiffres marqués `selon le fournisseur`, `estimation` ou `analyse` doivent être lus comme tels. Les noms d'entreprises et de produits appartiennent à leurs propriétaires et ne sont utilisés qu'à des fins d'identification ; aucune entreprise n'a relu ni parrainé ce rapport.

© 2026 Tan Shuai. Vous pouvez partager ce rapport sans modification, en citant l'auteur.

Sources

Toutes consultées entre le 3 et le 12 novembre 2026. Regroupées par chapitre.

Trafic, appareils et périmètre

- Ericsson, *Mobility Report*, juin 2026 (la vidéo représente environ 75 % du trafic de données mobiles fin 2025) : https://www.ericsson.com/en/reports-and-papers/mobility-report/reports/june-2026 ; résumé : 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, livraisons mondiales de smartphones, T4 et année complète 2025 (1,26 milliard), 13 janvier 2026 : https://www.idc.com/resource-center/press-releases/q425mobilephonetop5/
- TrendForce, prévision de livraisons de téléviseurs pour 2025 (195,71 millions) : 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, prévision 2025 pour les téléviseurs (208,7 millions) : https://iconnect007.com/article/145954/omdia-forecasts-flat-tv-shipments-in-2025-but-mini-led-poised-for-explosive-growth-at-829/145951/ein

Lecture et streaming

- Netflix Technology Blog, « AV1 — Now Powering 30% of Netflix Streaming », décembre 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 mai 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 » (part de l'AV1 dans le catalogue de 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 » (exploitation commerciale du DTV+, juin 2026) : https://mpegh.com/brazil/ ; CartaCapital sur les récepteurs intégrés aux téléviseurs : https://www.cartacapital.com.br/?p=628337 ; Olhar Digital sur le VVC dans la 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 octobre 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 octobre 2025 : https://atscaleconference.com/how-meta-deployed-super-resolution-at-scale-to-transform-video-quality/
- Samsung, points forts des téléviseurs 8K (NQ8 AI Gen3, 768 réseaux de neurones) : https://www.samsung.com/us/tvs/8k-tv/highlights/

Photographie et images

- 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 et Visionary.ai, ISP entièrement fondé sur l'IA, 5 janvier 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/ ; entretien de PetaPixel : https://petapixel.com/2026/01/18/a-look-at-the-worlds-first-full-ai-based-image-signal-processor
- Comité JPEG, communiqué de presse de la 103e réunion (gains, complexité et implémentations sur smartphone de JPEG AI) : https://jpeg.org/items/20240507_press.html
- G. J. Sullivan, introduction à l'ISO/IEC JTC 1/SC 29, avril 2026 (statut de JPEG AI ; chiffres d'ECM et de 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/
- Groupe de travail IEEE 1857.11 (codage d'image par réseaux de neurones) : https://sagroups.ieee.org/fvc
- Chrome 155 et JPEG XL, 6 octobre 2026 : https://xenospectrum.com/en/chrome-155-jpeg-xl-rust/

Appels et réunions

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

- Notes de version d'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 juin 2024 : https://engineering.fb.com/2024/06/13/web/mlow-metas-low-bitrate-audio-codec/
- Microsoft, PC Copilot+ (NPU de 40 TOPS et plus) : https://www.microsoft.com/en-us/windows/copilot-plus-pcs ; Apple, annonce du M4 (Neural Engine de 38 TOPS) : https://www.apple.com/newsroom/2024/05/apple-introduces-m4-chip/

IA générative

- Kyutai, fiche du modèle Mimi (12,5 Hz, 1,1 kbps) : https://huggingface.co/kyutai/mimi
- Lightricks, « LTX-Video: Realtime Video Latent Diffusion » (VAE vidéo 1:192) : https://arxiv.org/abs/2501.00103
- MPEG, communiqué de presse de la 153e réunion (Video Coding for Machines au stade DIS) : https://www.mpeg.org/wp-content/uploads/2026/03/press_release_153.pdf

Jeux et graphisme

- NVIDIA, adoption de DLSS 4 : lancement et Computex 2025 (75 et plus de 125) : https://www.nvidia.com/en-us/geforce/news/125-dlss-4-multi-frame-gen-games-more-announced-computex-2025/ ; Gamescom 2025 (plus de 175) : https://www.nvidia.com/en-us/geforce/news/gamescom-2025-dlss-4-full-ray-tracing-game-announcements/ ; CES 2026 (plus de 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 », février 2026 : https://www.ubisoft.com/en-us/studio/laforge/news/415slmB3ZGzv8d2LhO9QT/shipping-neural-texture-compression-in-assassins-creed-mirage
- TweakTown sur la démo de textures neuronales de NVIDIA au GTC 2026 (de 6,5 GB à 970 MB) : https://www.tweaktown.com/news/110886/nvidias-neural-texture-compression-cuts-vram-usage-from-6-5gb-down-to-970mb/index.html
- Blog Microsoft DirectX, Shader Model 6.9 en version finale, 26 février 2026 : https://devblogs.microsoft.com/directx/shader-model-6-9-retail-and-more/ ; sur les vecteurs coopératifs : https://devblogs.microsoft.com/directx/shader-model-6-9-and-the-future-of-cooperative-vector/ ; projets pour la GDC 2026 : https://windowsreport.com/microsoft-showcases-neural-rendering-and-dxr-2-0-plans-for-directx-at-gdc-2026/

Liaisons spécialisées

- Global Trade Alert sur l'annonce chinoise n° 49 de 2026 : 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 (seuils des plateformes terrestres sans pilote) : 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 ajoute les drones et composants de fabrication étrangère à la Covered List, 22 décembre 2025 : https://dronelife.com/2025/12/22/fcc-adds-foreign-made-drones-and-components-to-covered-list-citing-national-security-risks/

Silicium

- 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 ; page de la puce au MIT (78 mW en 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 ; documents complémentaires (complexité par module) : https://openaccess.thecvf.com/content/CVPR2025/supplemental/Jia_Towards_Practical_Real-Time_CVPR_2025_supplemental.pdf
- NNVC et la bibliothèque d'inférence en entiers SADL : https://arxiv.org/abs/2309.05846
- Tencent Canghai V2, 27 mai 2026 (Shanghai Securities News via 10jqka) : https://stock.10jqka.com.cn/20260527/c677016020.shtml ; synthèse des fonctionnalités : https://www.besthub.dev/articles/tencent-s-canghai-v2-chip-enters-mass-production-what-the-new-capabilities-mean-for-video-encoding-bcfc18dbfa29
- Chips&Media, IP de décodeur AV2, 12 mai 2026 : https://www.design-reuse.com/news/202530501-chips-media-completes-development-of-next-gen-av2-hw-decoder-ip/ ; licence AV2, 17 juin 2026 : https://www.edge-ai-vision.com/?p=58270
- Allegro DVT, IP de décodeur AV2, 16 juin 2026 : https://www.edge-ai-vision.com/2026/06/allegro-dvts-pulsar-decoder-ip-adds-support-for-av2-video-codec/
- VeriSilicon, VC9800D avec AV2, 10 juin 2026 : https://www.businesswirechina.com/en/news/63176.html

Normes et brevets

- MPEG, communiqué de presse de la 155e réunion (appel à propositions conjoint au-delà de VVC), 5 août 2026 : https://www.mpeg.org/wp-content/uploads/2026/08/ag3w254SN26474Press-Release-MPEG155.pdf
- MPEG, communiqué de presse de la 152e réunion (résultats de l'appel à preuves ; éditions de VVC et de VSEI), 30 octobre 2025 : https://www.mpeg.org/wp-content/uploads/2025/11/Press-Release-MPEG152.pdf ; chronique MPEG de l'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, cartographie des brevets post-VVC (H.267) : https://greyb.com/blog/exclusive-video-codec-landscape/ ; sur les licences AV2 et la vente du portefeuille d'Alibaba : https://greyb.com/blog/av2-licensing/
- MLex, « Fragmented codec licensing exposes need for clearer rules for Chinese tech majors », 27 mars 2026 : https://www.mlex.com/mlex/articles/2458402/fragmented-codec-licensing-exposes-need-for-clearer-rules-for-chinese-tech-majors
- Access Advance, taux de redevances MCBA et VVC Advance au 1er juillet 2026 : https://accessadvance.com/wp-content/uploads/2025/07/MCBA-Program-Overview-September-2026.pdf

Argent et entreprises

- InterDigital rachète Deep Render, 30 octobre 2025 : https://www.advanced-television.com/2025/10/30/interdigital-acquires-deep-render
- TechCrunch, « Apple acquired a startup using AI to compress videos », 27 mars 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/

COMPRESSION APPRISE · LECTURE × SILICIUM

Le codage neuronal en 2026

Il se déploie autour du codec, pas à sa place

Chaque chiffre clé de ce rapport est accompagné de sa source et de sa date. Lorsque l'auteur calcule lui-même un chiffre, celui-ci est marqué comme estimation ; lorsque seul le vendeur le communique, il porte la mention « selon le fournisseur ».

Tan Shuai

Technologue et fondateur · Chercheur et développeur en IA

tanshuai.com · i@tanshuai.com

ÉDITION PUBLIQUE · Données au 12 novembre 2026