리서치 노트 · 2026년 11월

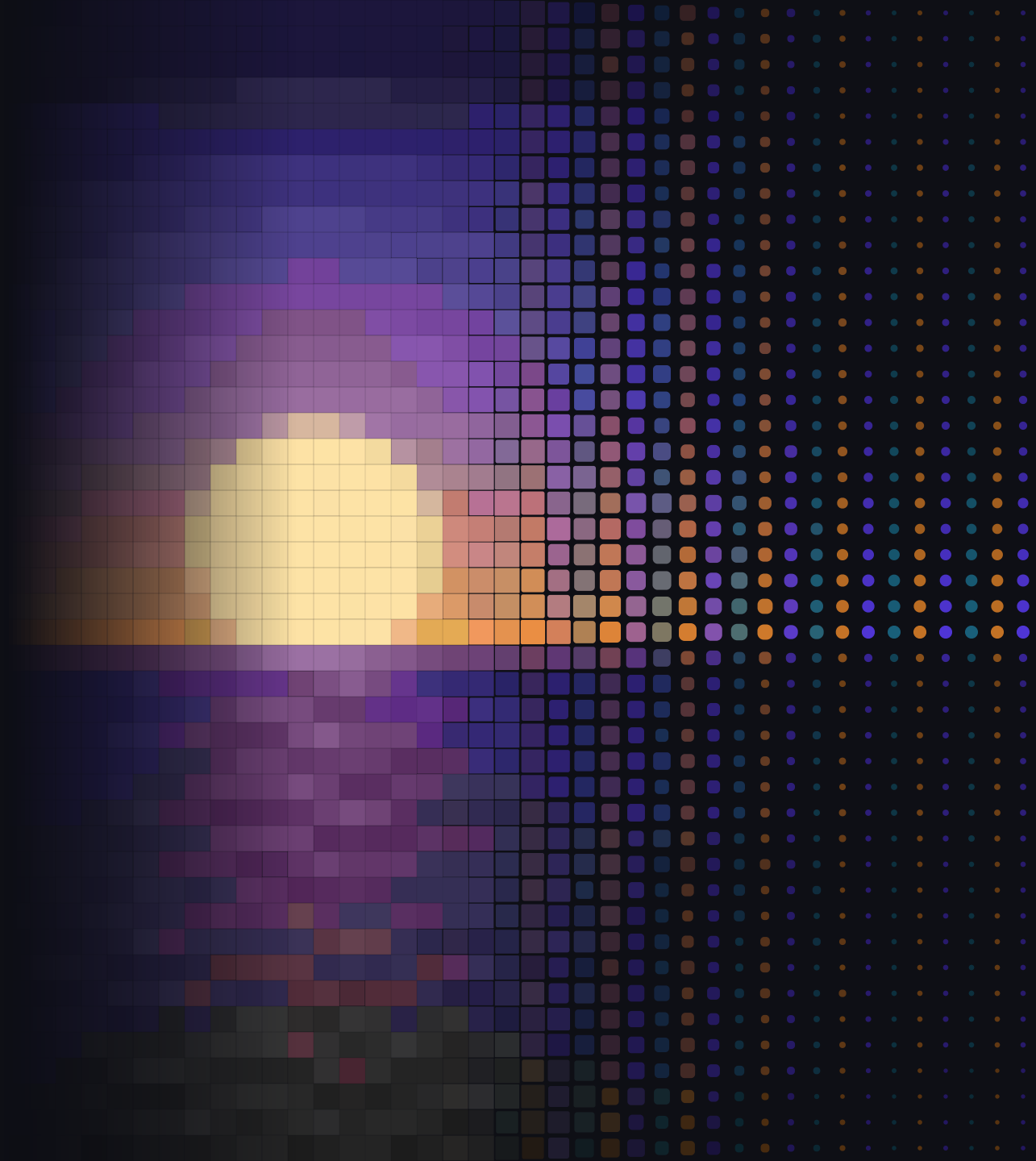
2026년 신경망 코딩

코덱 대체가 아니라 코덱 주변에서

재생 · 사진 · 통화 · 생성형 AI · 게임 · 실리콘

탄 슈아이

tanshuai.com · i@tanshuai.com



이미 주류지만, 코덱 주변에서

01 신경망은 표준 코덱의 양 끝에서 이미 주류다

카메라 ISP, 인코딩 측 다운스케일링, 재생 시 초해상도. 그 사이의 비트스트림은 여전히 AV1, HEVC, AVC, VVC이며 고정 기능 하드웨어가 디코딩한다.

02 엔드투엔드 신경망 코덱은 폐쇄 루프에서만 상용화된다

통화, 생성형 AI 토크나이저, AAA 게임 한 편. 모두 한 회사가 양 끝을 통제하고 표준 코덱으로 폴백할 수 있다.

03 대중 기기의 신경망 비트스트림은 H.267을 기다린다

버전 1은 2029년 말 예정이고, 첫 칩은 2031년경, 대중화는 2030년대 중반이다 추정.

≈75%

모바일 데이터 트래픽 중 비디오 비중(2025년 말)

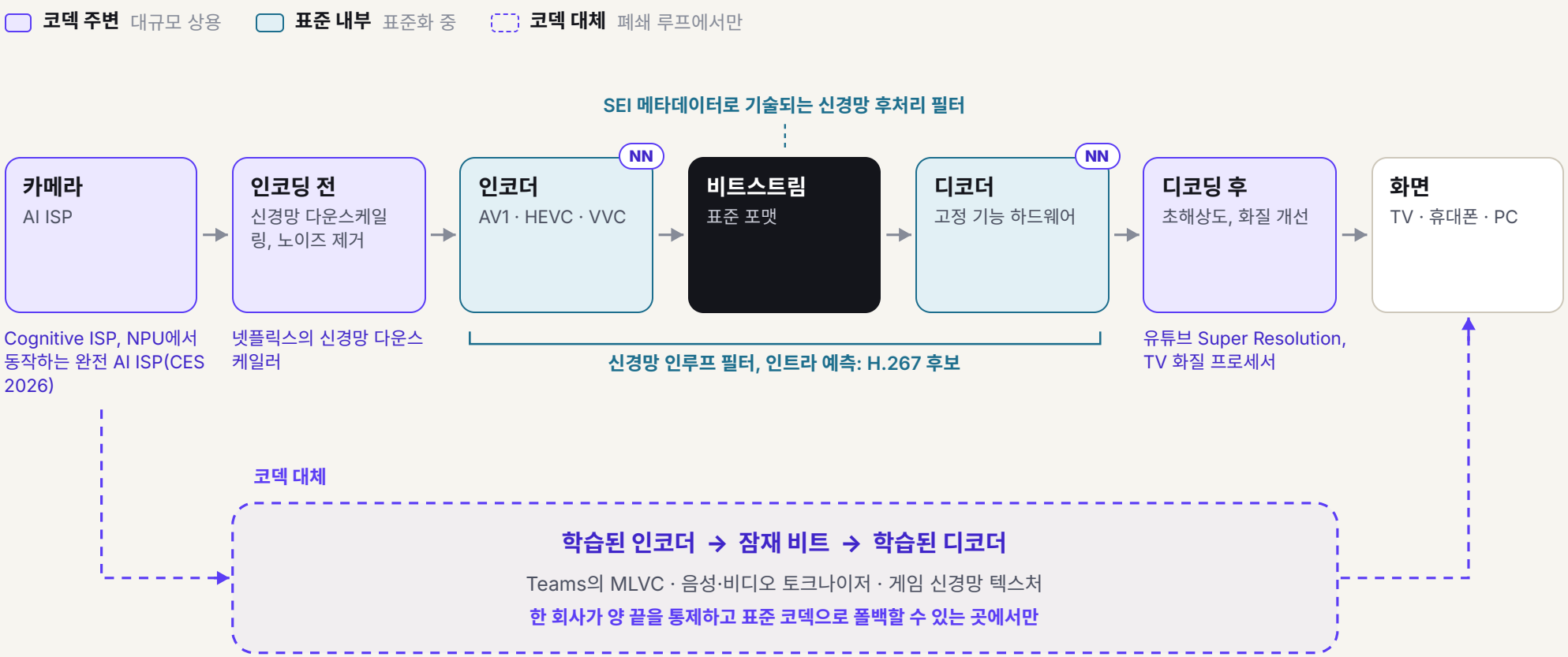
12억 6천만 대

2025년 스마트폰 출하량. 모두 고정 기능 하드웨어로 비디오를 디코딩한다

≈12배

유일하게 공개된 신경망 디코더 칩 설계의 픽셀당 에너지, 하드웨어 HEVC 디코더 대비

신경망이 코덱에 들어가는 세 가지 방식



코덱 주변은 누구의 허락도 필요 없다. 표준 내부는 표준이 필요하다. 코덱 대체는 양 끝을 모두 통제해야 한다.

상용화 현황

2026년 신경망 코딩이 상용화된 곳

	표준 코덱 주변 전처리, 화질 개선, 업스케일링	엔드투엔드 신경망 코덱 학습된 인코더와 디코더	표준 내부 표준 비트스트림이나 API 안의 신경망 도구
01 재생과 스트리밍	● 대규모 상용 넷플릭스 다운스케일러, 유튜브·메타 초해상도, TV AI 업스케일러	● 미도입 신경망 비디오 포맷을 송출하는 스트리밍 서비스는 없음. TV와 휴대폰은 표준 코덱을 하드웨어로 디코딩	● 표준화 중 H.267 버전 1은 2029년 말 예정, 신경망 도구도 후보에 포함
02 사진과 이미징	● 대규모 상용 플래그십 휴대폰의 AI ISP, CES 2026에서 IP로 공개된 완전 신경망 ISP	● 시연·시범 JPEG AI는 ISO 표준: 휴대폰 시연은 있으나 제품은 없음	● 미도입 확산 중인 포맷은 전통 방식: Chrome 155에서 JPEG XL 기본 활성화
03 통화와 회의	● 대규모 상용 수신 측 초해상도와 신경망 잡음 억제	● 대규모 상용 2021년부터 Teams 일대일 통화에 신경망 오디오, Teams에 MLVC 비디오 확대 적용 중	● 제한적 상용 Opus 1.5에 선택형·하위 호환 ML 도구 추가
04 생성형 AI	—	● 대규모 상용 모든 비디오·음성 모델이 학습된 토크나이저를 사용 (예: 1:192 비디오 VAE)	—
05 게임과 그래픽	● 대규모 상용 신경망 업스케일링·프레임 생성: DLSS 4만 해도 250개 이상의 게임·앱	● 제한적 상용 AAA 게임 한 편에 신경망 텍스처 탑재, 엔비디아 SDK는 베타	● 표준화 중 DirectX에 셰이더용 행렬 연산 추가 중(프리뷰)
06 특수 링크: 위성, 로봇	● 시연·시범 저비트레이트 영상의 서버 측 복원	● 시연·시범 손실에 강한 신경망 코덱, 연구·시범 단계	—

행은 각 영역이 닿는 사람 수 순으로 배열했다.

양 끝은 신경망, 가운데는 표준

- **인코더 쪽**

  넷플릭스의 신경망 다운스케일러는 2022년부터 어떤 기기도 바꾸지 않고 운영 중이다. 넷플릭스 시험에서 시청자의 약 77%가 이를 선호했다 업체 주장.

- **화면 쪽**

  2025년 10월부터 TV에서 유튜브 Super Resolution이 동작한다. 메타는 서버와 휴대폰 양쪽에서 초해상도를 돌린다.

- **TV 안**

  삼성전자의 NQ8 AI Gen3 프로세서는 업스케일링용 신경망 768개를 내세운다.

비트스트림은 표준 그대로

30%

넷플릭스 시청 중 AV1 비중(2025년 12월)

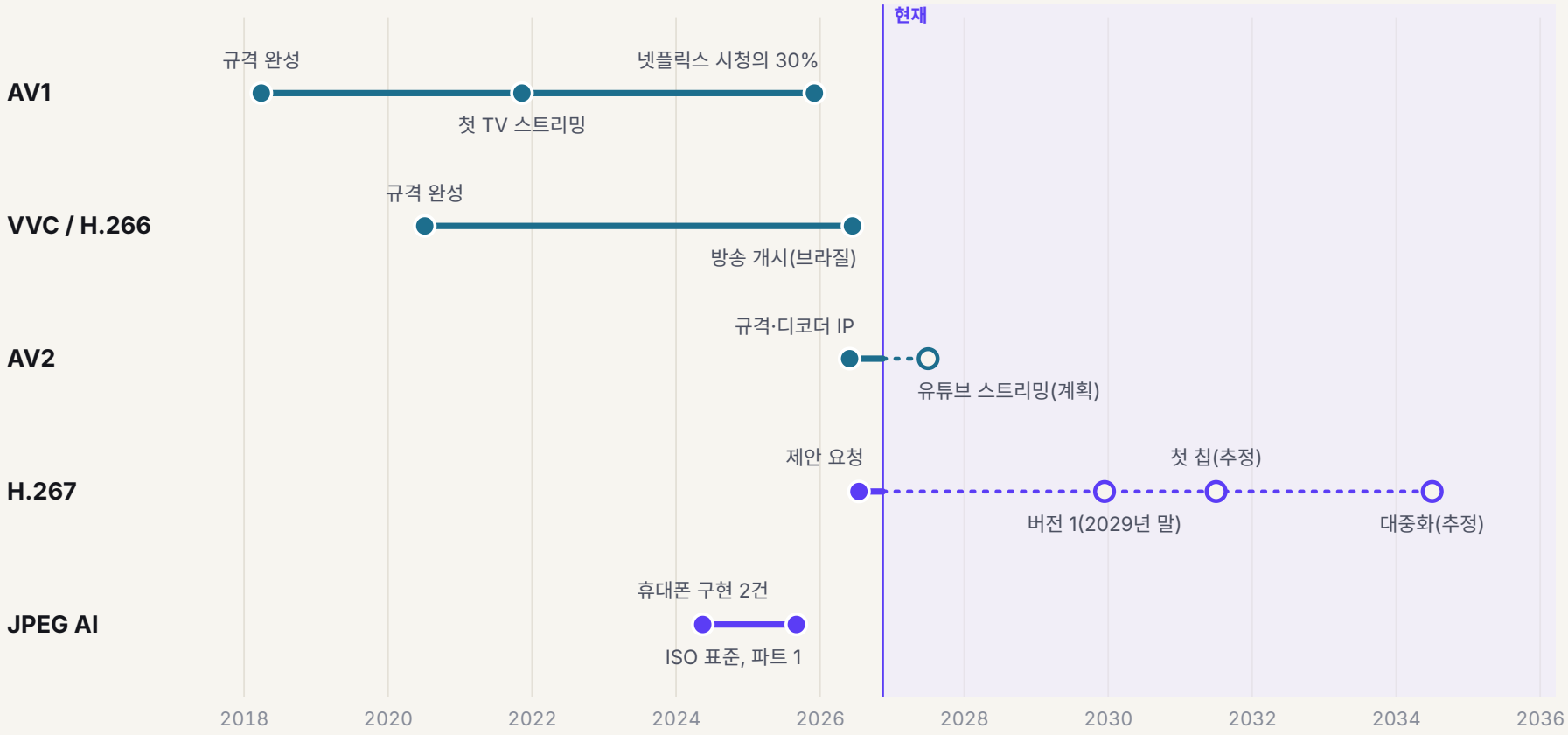
90%+

시청 시간 기준 유튜브 카탈로그 중 AV1 제공 비중(2026년 3월)

3주

AV2 버전 1.0.0(2026년 5월 28일)에서 IP 업체 세 곳의 디코더 IP 고객 확보까지

표준 완성에서 대중 기기 보급까지 10년 가까이 걸린다



AV1은 규격 완성(2018년)에서 넷플릭스 시청의 30%(2025년 12월)에 이르기까지 8년 가까이 걸렸다. 속이 빈 표시는 목표치 또는 필자의 추정이다.

신경망 도구가 모든 화면에 닿을지는 H.267이 결정한다

2025년 10월
증거 제출 요청: 16개 기업·연구소의 응답 5건, 신경망 이득은 8K HDR에서 가장 뚜렷

2026년 7월
제네바에서 제안 요청 발표

2027년 1월
브리즈번에서 응답 평가

2027년 하반기
첫 테스트 모델

2029년 말
표준 버전 1

≈28%
전통 방식 ECM 트랙의 VVC 대비 절감률

6–14%
픽셀당 1만~2만 2천 이상의 MAC을 쓰는 신경망 인루프 필터의 절감률

SEI
VVC 스트림은 디코더가 적용할 신경망 후처리 필터를 이미 지정할 수 있다

필자의 판단: 절감되는 비트가 디코더 비용을 정당화하는 곳에만 신경망 도구 몇 개를 얹은 전통 방식 코어가 될 것이다 분석.

카메라 안은 신경망, 웹은 전통 방식

- **카메라**

  퀄컴의 Cognitive ISP는 2023년부터 휴대폰에서 실시간 세그멘테이션을 수행해 왔다. NPU에서 동작하는 완전한 AI ISP가 CES 2026에서 공개됐다.

- **JPEG AI**

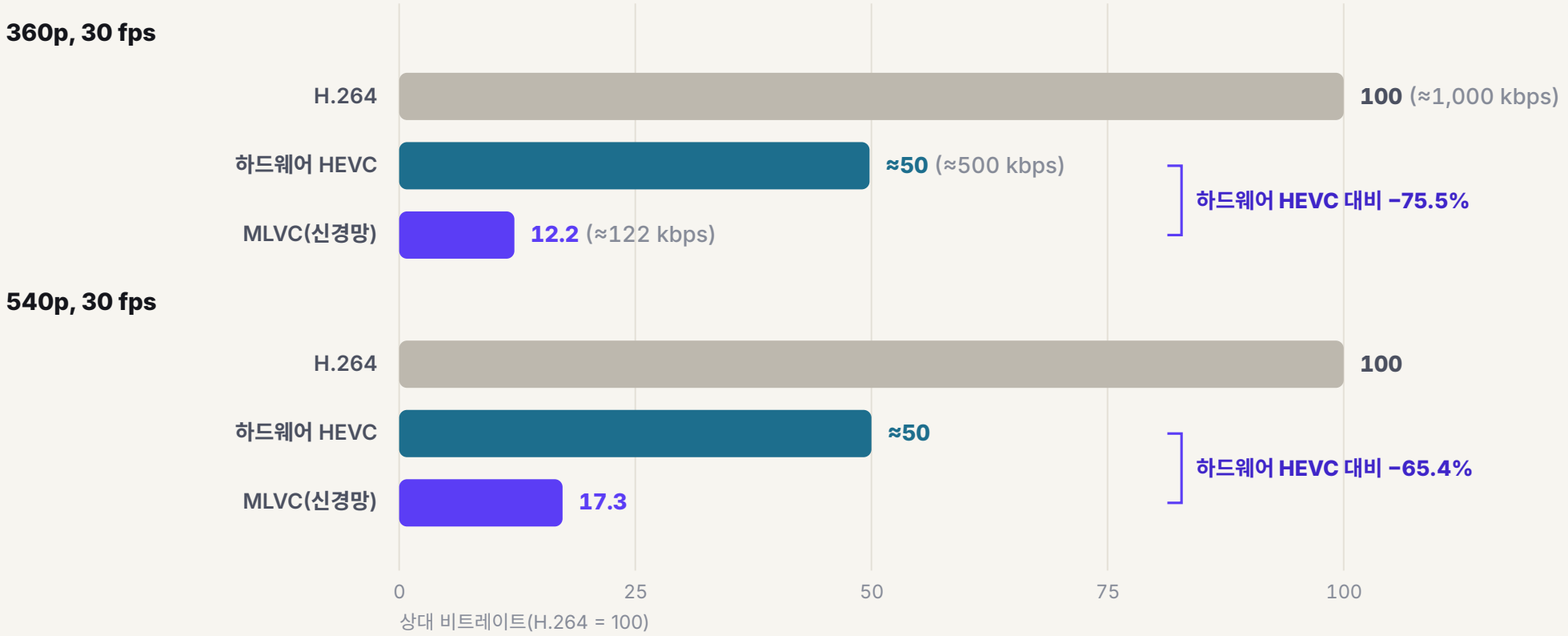
  2025년 발행. 픽셀당 7천~21만 6천 MAC으로 VVC 인트라 대비 12.5–27.9%의 이득. 이를 기본으로 쓰는 제품은 찾지 못했다.

- **웹**

  Chrome 155는 2026년 10월 6일 Rust로 작성한 메모리 안전 디코더와 함께 JPEG XL을 기본 활성화했다.

> **어디서나 열려야 하는 포맷은 전통 방식으로 남고, 신경망 코딩은 한 회사의 파이프라인 안에서 이긴다.**

엔드투엔드 신경망 코덱이 상용화된 곳은 통화다



Satin은 2021년부터 Teams와 Skype의 모든 일대일 통화에 쓰여 왔다. 2026년 7월 24일부터 MIT 라이선스의 오픈소스인 MLVC는 Teams P2P 통화에 최대 540p30으로 확대 적용 중이며 NPU를 절반 이하로만 쓴다 업체 주장.

모든 플랫폼이 신경망을 택하지는 않았고, 그 이유가 시사적이다

표준 내부, 선택 사항

Opus 1.5와 1.6

신경망 음질 개선, 패킷 손실 은닉, 딥 리던던시, 대역폭 확장. 모두 기본값은 꺼져 있고 2012년 표준과 완전히 호환된다.

의도적인 전통 방식

메타 MLow(2024)

메타 통화의 20% 이상이 ARMv7 휴대폰에서, 하루 수천만 건의 WhatsApp 통화가 10년 넘은 휴대폰에서 이뤄진다. MLow는 6 kbps에서 Opus의 약 2배 품질을 내면서 연산량은 10% 적다.

결정하는 것은 이미 보급된 기기의 복잡도 예산이다. 신경망 코덱은 기기가 새것이고 한 업체가 양 끝을 통제하는 곳에서 먼저 상용화된다.

아무도 코덱이라 부르지 않는 최대 규모의 배포

- **어디에나 있는 학습된 코덱**

  모든 음성·비디오 모델은 학습된 코덱의 토큰이나 잠재값을 읽고 쓴다.

- **배포용 포맷이 아니다**

  모델마다 고유한 잠재 공간이 있다. 레이트 제어도, 적합성 시험도, 두 번째 디코더도 없다.

- **비용 항목**

  생성형 비디오를 서비스하는 쪽에서는 압축률이 높은 잠재 표현일수록 생성 비용이 비례 이상으로 줄어든다 분석.

두 가지 사례

12.5 Hz · 1.1 kbps

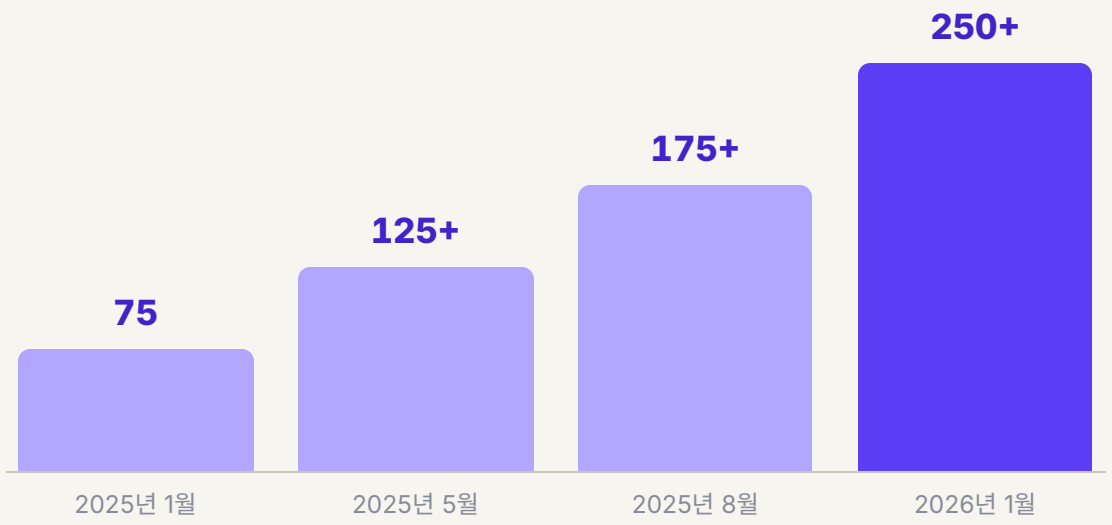
Kyutai Mimi의 음성 토큰. 어떤 전화 코덱보다 낮다

1:192

LTX-Video의 비디오 VAE: 잠재 토큰 하나가 8프레임에 걸친 32×32 픽셀을 나타낸다

신경망 복원은 주류, 신경망 에셋은 시작 단계

DLSS 4 Multi Frame Generation 지원 게임·앱



신경망 텍스처 압축

엔비디아 시연 장면, GTC 2026

텍스처 메모리 6.5 GB → 0.97 GB

메모리 85% 절감

Assassin's Creed Mirage(출시작)

일부 에셋에만 적용, 블록 포맷 대비

압축률 ≈30% 향상

측정 기준이 달라 직접 비교할 수 없음

DirectX Linear Algebra가 2026년 4월 퍼블릭 프리뷰에 들어갔다. 어느 업체의 GPU에서든 셰이더 안의 신경망 디코딩을 하나의 코드 경로로 처리할 수 있게 된다.

위성, 드론, 로봇: 가장 잘 맞지만 시장은 작다

- **잘 맞는 이유**

  한 소유자가 양 끝을 통제하고, 대역폭이 부족하며, 콘텐츠 범위가 좁아 학습시키기 쉽다.

- **중국**

  11월 18일 시행되는 2026년 제49호 공고가 드론 수출 통제를 개정한다. 한 항목은 시맨틱 이미지 압축을 기준치로 쓴다.

- **미국**

  FCC는 2025년 12월 22일 외국산 드론과 그 핵심 부품을 Covered List에 올렸다.

> **수출 규정 준수 역량을 갖춘 전문 기업에는 탄탄한 틈새시장이지만, 주류 시장은 아니다.**

픽셀당 에너지가 10배 수준 더 든다

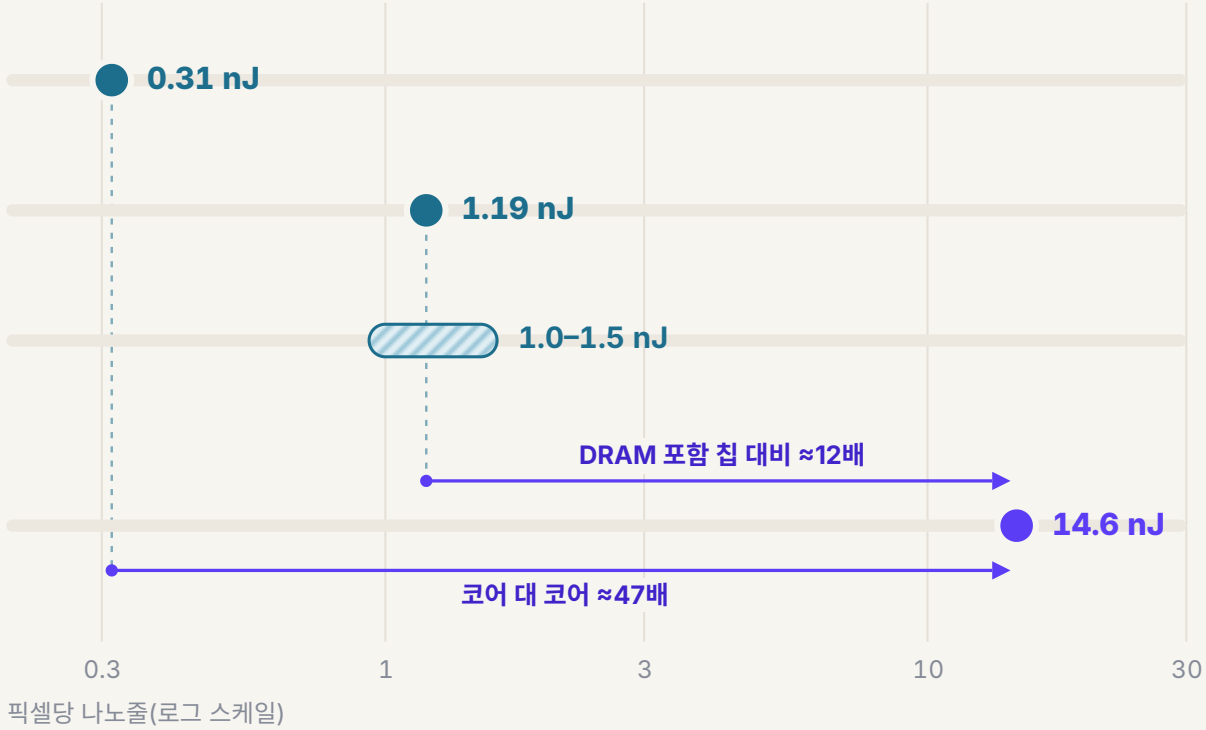
HEVC 디코더 칩, 코어
40 nm, 4K30, 실측

HEVC 디코더 칩, DRAM 포함
40 nm, 4K30, 코어 실측 + DRAM 시뮬레이션

VVC 디코더
28 nm, 4K60, 추산

신경망 비디오 디코더(NVCA)
28 nm, 1080p25, 합성 결과만, 가중치·DRAM 미포함

1080p30으로 환산하면 약 74 mW 대 908 mW다.



1080p30 기준으로 DRAM을 포함한 HEVC 칩은 약 74 mW, NVCA는 약 900 mW다. HEVC 칩은 더 오래된 공정이고 NVCA는 가중치와 DRAM을 포함하지 않았으므로, 이 비교는 신경망 설계에 유리하다.

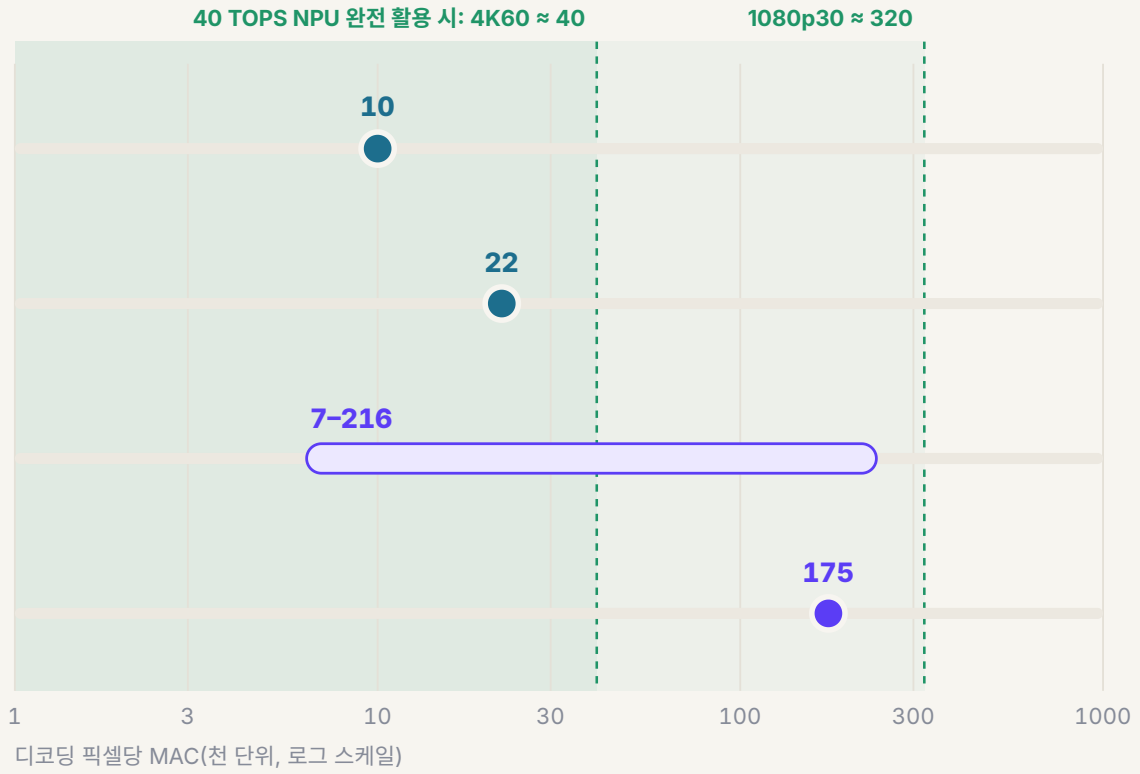
신경망 디코더가 NPU에 요구하는 것

신경망 인루프 필터, 초저복잡도
JVET 탐색 연구 · 비트 ≈6% 절감
1080p30에서 1.2 TOPS · 4K60에서 10

신경망 인루프 필터, 저복잡도
JVET 탐색 연구 · ≈8%
1080p30에서 2.7 TOPS · 4K60에서 22

JPEG AI 디코더(기본형~고성능형)
VVC 인트라 대비 12.5–27.9%

DCVC-RT, 신경망 비디오 디코더 전체
VVC 참조 인코더 대비 21%, 논문 조건
1080p30에서 22 TOPS · 4K60에서 174



DCVC-RT는 픽셀당 약 17만 5천 MAC이 필요하다. 완전 활용을 가정하면 1080p30에서 약 22 TOPS, 4K60에서 174 TOPS다 추정. 완전히 활용한 40 TOPS NPU가 4K60에서 감당할 수 있는 양은 약 4만 MAC이다.

메모리 트래픽과 비트 단위 일치가 넘기 어려운 한계다

DRAM이 지배적

MAC보다 메모리

MIT HEVC 칩에서 DRAM이 차지하는 몫은 시스템 에너지의 약 4분의 3, 즉 픽셀당 1.19 nJ 중 0.88 nJ에 이른다. 신경망 디코더는 프레임 사이에 큰 특징 맵을 유지하며, DCVC-RT 저자들은 속도를 제한한 것이 산술 연산이 아니라 메모리 입출력이었다고 밝혔다.

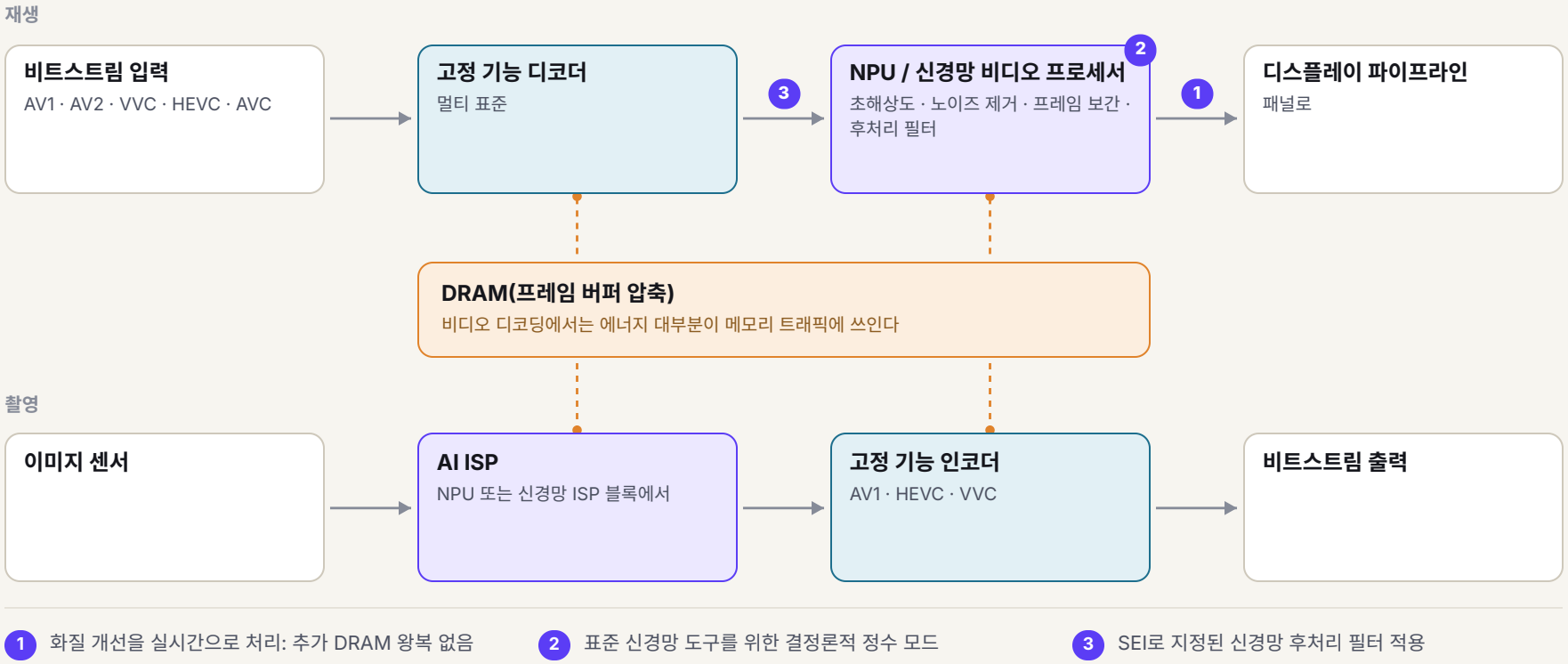
정확성

모든 디코더가 같은 비트를 계산해야 한다

두 기기가 확률을 다르게 계산하면 엔트로피 디코딩이 깨진다. 해법은 정수 모델(DCVC-RT), 공유 스케일 하이퍼프라이어(MLVC), 16비트 정수 추론(JVET의 SADL), 적합성 검증 지점(JPEG AI)이다.

칩 팀에게는 HEVC 정수 변환의 규율이 되풀이되는 셈이다. 표준이 산술 연산을 정확히 정의해야 한다.

상용화된 분업 구조: 고정 기능 코덱 + NPU



> 텐센트 Canghai: 1세대 칩 10만 개 이상이 배치되었고, 2세대는 H.266과 온칩 초해상도를 추가했다. 칩스앤미디어와 Allegro DVT는 코덱 IP와 신경망 비디오 IP를 나란히 판매한다.

역할별로 필자라면 무엇을 만들까

SoC·TV 칩 팀

- 재생해야 하는 모든 표준의 고정 기능 디코더, 지금은 AV2
- 디스플레이 해상도의 화질 개선에 맞춘 NPU, 추가 DRAM 왕복 없이
- 신경망 후처리 필터 SEI와 결정론적 정수 모드
- H.267 설계 착수는 2029년 이후

IP 업체

- 지금은 AV2·VVC 디코더
- 후처리 필터 가속기와 프레임 버퍼 압축
- 신경망 처리용 적합성 도구

비디오 플랫폼

- 인코더 쪽 신경망 전처리가 가장 먼저 투자를 회수한다
- 기기가 허용하는 곳에서는 클라이언트 측 초해상도
- 엔드투엔드 신경망 코덱은 직접 통제하는 폐쇄 루프에서만

신경망 도구가 라이선스 대상이 되는 곳

표준	신경망 요소	현황(2026년 11월)
H.267(가칭)	미정: 전통 방식(ECM)과 신경망 도구 모두 후보	2027년 1월 응답 평가, 2029년 말 버전 1
VVC·HEVC(VSEI 포함)	SEI로 기술되는 신경망 후처리 필터, 디코더에는 선택 사항	시행 중, 2025년 10월 확장
AV2	전통 방식, 로열티 무료	2026년 5월 28일 버전 1.0.0, 유튜브 웹 스트리밍은 2027년 초 계획
JPEG AI	엔드투엔드 학습 기반 이미지 코덱	2025년 파트 1 발행, 제품의 기본 지원은 확인되지 않음
IEEE 1857.11	NPU용으로 설계된 신경망 이미지 코딩	2024년 12월 발행
Video Coding for Machines	시청이 아닌 기계 분석을 위한 코딩	2026년 초 국제 표준안
Opus ML 확장	선택 사항, 하위 호환	libopus 1.5와 1.6에 포함

645

바이트댄스가 보유한 VVC 이후 세대 특허 패밀리(GreyB, 2025)

0.50달러 · 1.00–1.50달러

VVC 풀 요율, 휴대폰당·4K TV당, Region 1

0달러

AOMedia 회원사가 AV1·AV2에 매기는 로열티

누가 돈을 버는가

돈은 표준과 플랫폼을 따라간다

현재 수익

플랫폼

절감분을 가져간다

넷플릭스 신경망 다운스케일러, 유튜브·메타 초해상도, 마이크로소프트는 MLVC를 MIT 라이선스로 공개

현재 수익

기기·칩 제조사

차별화를 판다

카메라 ISP, TV 업스케일링 프로세서, 텐센트가 자체 개발한 Canghai 코덱 칩

현재 수익

IP 업체

새 표준마다 라이선스한다

규격 확정 3주 만에 칩스앤미디어, Allegro DVT, 베리실리콘이 AV2 디코더 고객 확보

현재 수익

GPU 업체

플랫폼 종속을 만든다

250개 이상의 게임·앱에 DLSS 4, 신경망 텍스처 SDK는 베타

기기당

특허권자와 특허 풀

기기당 로열티를 걷는다

VVC: 휴대폰당 0.50달러, 4K TV당 1.00–1.50달러, 2029년 이후 H.267 도구

인수로 엑시트

신경망 코덱 스타트업

인수됐다

WaveOne → 애플(2023), iSIZE → 소니(2023), Deep Render → 인터디지털(2025)

> 가장 잘 알려진 신경망 코덱 스타트업 세 곳, WaveOne, iSIZE, Deep Render는 각각 애플, 소니, 인터디지털에 인수됐다.

기업이 기회를 찾을 수 있는 곳

여섯 가지 사업 기회와 각각의 위험

사업 기회	구매자	지금인 이유	주요 위험
인코더 쪽 신경망 전처리	스트리밍·사용자 동영상 플랫폼, 방송사	기기 변경 불필요, 넷플릭스에서 검증	대형 플랫폼의 자체 개발
디스플레이 측 화질 개선 IP와 튜닝	TV·셋톱·휴대폰 칩 제조사	기기에 이미 NPU 탑재	범용화, 사내 팀
폐쇄 루프 신경망 코덱	기업 통신·산업용·텔레프레즌스 업체	MLVC가 작동을 입증했고 코드도 공개	양 끝을 통제해야 함, NPU 파편화
코덱 + NPU 실리콘 또는 IP	SoC 제조사	AV2 사이클 시작	강력한 기존 강자
평가, 적합성, 결정론적 추론	신경망 도구를 출시하는 모든 기업	비트 단위 일치는 필수 요건	묶어 팔지 않으면 작은 시장
H.267 표준 기고	라이선서, 대형 플랫폼	2027년 1월 응답 평가	길고, 비싸고, 불확실

흔히 듣게 될 주장

네 가지를 밝히기 전까지는 주장을 에누리해 들어라

주장	근거가 말하는 것
“신경망 코덱은 이미 VVC를 앞선다.”	실험실에서는 그렇다. DCVC-RT는 VVC 참조 인코더 대비 21%를 절감하지만, 디코더 비용은 픽셀당 약 17만 5천 MAC이다.
“AI 코덱이 몇 년 안에 H.26x와 AV1을 대체할 것이다.”	재생에서는 아니다. H.267 버전 1은 2029년 말, 칩은 2031년경이다 추정. 지금 도입되는 코덱인 AV2는 전통 방식이다.
“MLVC는 대역폭을 88% 줄인다.”	360p에서 H.264 대비 수치다. 하드웨어 HEVC 대비로는 마이크로소프트 자체 테스트셋에서 360p 75.5%, 540p 65.4%다.
“신경망 텍스처는 비디오 메모리를 85–96% 줄인다.”	시연 장면에서의 이야기다. 출시된 게임은 일부 에셋에서만 약 30%를 얻었다.
“JPEG AI는 ISO 표준이니 카메라와 브라우저가 채택할 것이다.”	제품에서 기본 지원은 확인되지 않았다. 웹이 채택한 것은 전통 방식 포맷인 JPEG XL이다.

기준, 지표, 해상도, 디코딩 비용을 물어라. 네 가지를 모두 밝히지 못하는 주장은 제품이 아니라 연구 결과다.

주목할 점

주목할 점과 열린 질문

이정표

2027년 1월	H.267 응답 평가: 몇 건이 신경망 도구에 의존하는지, 픽셀당 비용은 얼마인지
2027년 초	유튜브의 첫 AV2 웹 스트리밍, 발표된 SoC의 AV2 디코더
2027년 하반기	H.267의 첫 테스트 모델
2027년	Teams P2P 통화 너머로의 MLVC 확산
2027–2028년	JPEG AI를 기본으로 쓰는 첫 제품
2029년 말	H.267 버전 1, 실리콘 사이클 시작

열린 질문

01 대규모 정확성: 업체마다 다른 NPU에서 비트 단위로 일치하는 정수 디코딩을 하려면 압축 효율을 얼마나 잃는가?

02 MAC보다 메모리: 신경망 후처리의 픽셀당 DRAM 트래픽

03 H.267에서 신경망 도구가 제값을 못 하게 되는 픽셀당 비용

04 통화 외에 어떤 폐쇄 루프가 엔드투엔드 신경망 코덱을 먼저 채택할 것인가

05 디코더가 디테일을 생성할 때 충실도를 어떻게 측정할 것인가

기술 교류와 기술 협력을 통해 함께 탐구할 수 있는 분야들이다.

i@tanshuai.com · tanshuai.com