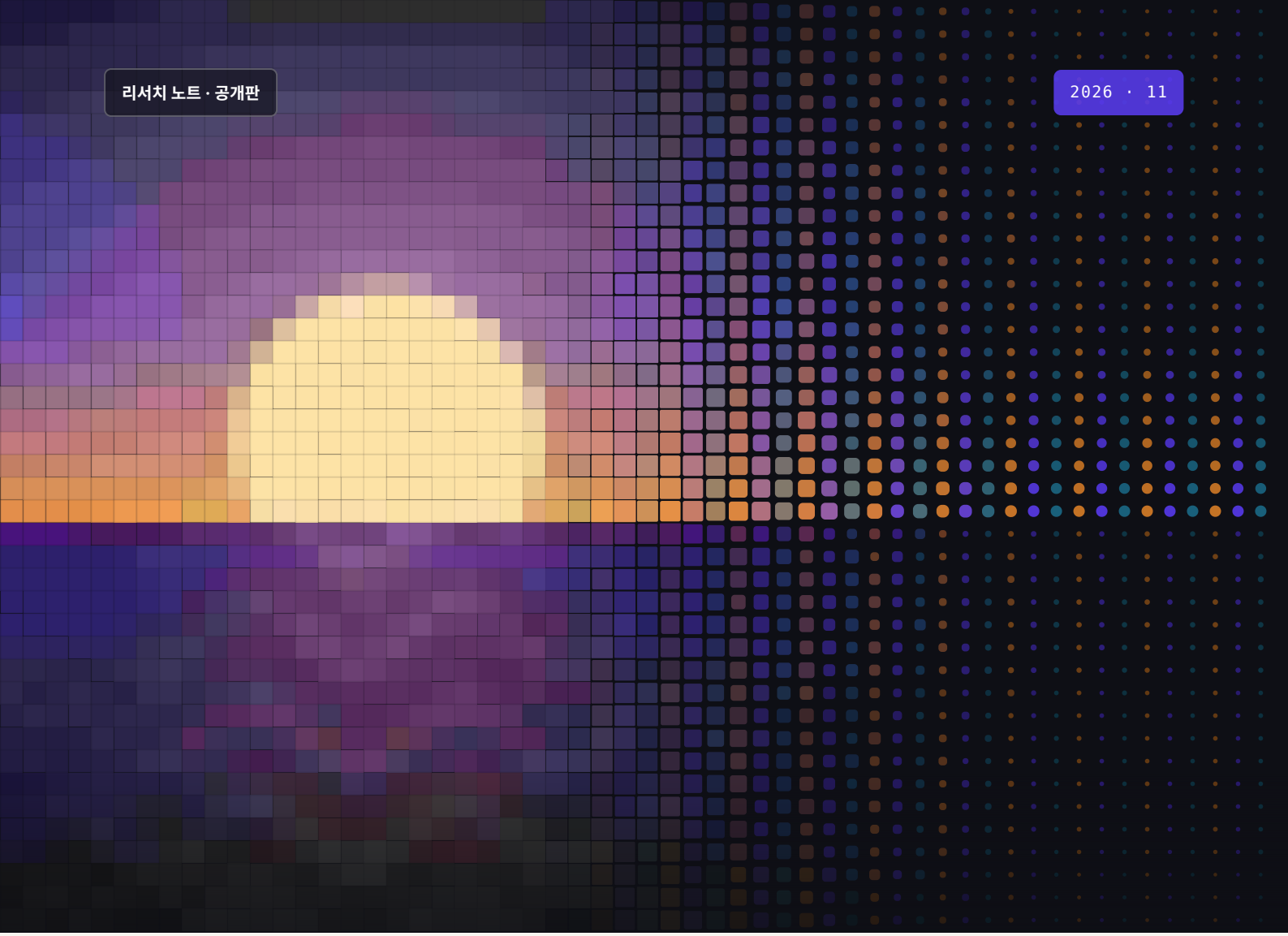


학습 기반 압축 · 재생 × 실리콘

2026년 신경망 코딩

코덱을 대체하지 않고 코덱 주변에서 상용화된다

재생 · 사진 · 통화 · 생성형 AI · 게임 · 실리콘

> 결론
>
> **신경망은 이미 전 세계 비디오를 실어 나르는 코덱의 양 끝에서 작동한다. 엔드투엔드 신경망 코덱은 폐쇄 루프에서만 상용화되며, 대중 기기는 H.267을 기다린다.**

탄 슈아이

기술 전문가이자 창업가 · AI 연구자이자 개발자

tanshuai.com · i@tanshuai.com

2026년 11월 12일 기준 데이터
금액은 미국 달러 기준

목차

목차

◆ **핵심 요약** 04
이미 주류지만, 코덱 주변에서

01 **범위와 방법** 07
‘신경망 코딩’이 여기서 뜻하는 것과 살펴본 방법

02 **재생과 스트리밍** 08
재생과 스트리밍: 양 끝은 신경망, 가운데는 표준

03 **사진과 이미지** 11
사진과 이미지: 카메라 안은 신경망, 웹은 전통 방식

04 **통화와 회의** 12
통화와 회의: 엔드투엔드 신경망 코덱이 실제로 상용화된 곳

05 **생성형 AI** 14
생성형 AI: 아무도 코덱이라 부르지 않는 최대 규모의 배포

06 **게임과 그래픽** 15
게임과 그래픽: 신경망 복원은 주류, 신경망 에셋은 시작 단계

07 **특수 링크** 16
특수 링크: 위성, 드론, 로봇

08 **실리콘** 17
실리콘 관점: 에너지, 연산, 메모리, 정확성

09 **표준과 특허** 21
표준과 특허: 신경망 도구가 라이선스 대상이 되는 곳

10 **누가 돈을 버는가** 22
누가 돈을 버는가, 기업은 어디서 기회를 찾을 수 있는가

11 **주장과 근거** 23
흔히 듣게 될 주장과 근거가 말하는 것

12 **이정표** 24
2027년부터 2030년까지 주목할 점

13 **열린 질문** 25

◆ **저자 소개** 26
탄 슈아이

◆ **출처** 27

그림 목록

그림



그림 01	2026년 신경망 코딩이 상용화된 곳	06
그림 02	신경망이 코덱에 들어가는 세 가지 방식	07
그림 03	표준 완성에서 대중 기기 보급까지 10년 가까이 걸린다	10
그림 04	통화에서 신경망 코덱은 하드웨어 HEVC 대비 비트레이트를 65–76% 줄인다	13
그림 05	한 회사가 스택 전체를 통제하면 신경망 복원은 빠르게 확산된다	15
그림 06	디코딩 픽셀당 에너지: 고정 기능 디코더가 10배 수준 앞선다	17
그림 07	신경망 도구는 픽셀당 수만 MAC, 신경망 디코더 전체는 수십만 MAC이 든다	18
그림 08	상용화된 분업 구조: 비트스트림은 고정 기능 하드웨어, 픽셀은 신경망 엔진	20
그림 09	2026년 신경망 코딩으로 누가 돈을 버는가	22

본문의 태그

추정	인용한 데이터를 바탕으로 한 저자의 계산 또는 추론
업체 주장	제품을 판매하는 회사가 발표한 것으로, 독립적으로 검증되지 않음
분석	근거에 대한 저자의 해석

이 문서는 영어판을 번역한 것이며, 두 판의 내용이 다를 경우 영어판이 우선한다.

2026년 11월 12일 기준 데이터 · 금액은 미국 달러 기준

핵심 요약

이미 주류지만, 코덱 주변에서

결론

2026년 신경망 코딩은 이미 주류지만, 표준 코덱을 대체하지 않고 그 주변에서 작동한다.

≈75%
모바일 데이터 트래픽 중 비디오 비중, 2025년 말(에릭슨)

12억 6천만 대
2025년 스마트폰 출하량. 모두 고정 기능 하드웨어로 비디오를 디코딩한다

≈12배
유일하게 공개된 신경망 디코더 칩 설계의 픽셀당 에너지. DRAM을 포함해 실측한 HEVC 디코더 칩 대비

≈22 TOPS
1080p30에서 DCVC-RT급 신경망 디코더에 필요한 지속 연산 성능 추정

75.5%
통화 360p에서 MLVC가 하드웨어 HEVC보다 줄인 비트레이트. 마이크로소프트 자체 테스트 셋 기준 업체 주장

2029년 말
H.267 버전 1. 신경망 도구를 모든 화면에 실어 나를 가능성이 큰 첫 표준

상용화 현황

신경망은 이제 전 세계 비디오의 거의 전부를 실어 나르는 코덱의 양 끝에서 돌아간다. 입구에서는 영상을 다듬는다. 휴대폰 카메라의 AI 이미지 신호 처리기와 넷플릭스 인코딩 파이프라인의 신경망 다운스케일러가 그렇다. 출구에서는 영상을 복원한다. TV에서 동작하는 유튜브의 Super Resolution, 메타의 초해상도, 수백 개의 신경망을 내세우는 TV 프로세서가 그렇다. 그 사이의 비트스트림은 여전히 AV1, HEVC, AVC, VVC이며, 고정 기능 하드웨어가 디코딩한다. TV나 휴대폰에 신경망 비디오 포맷을 보내는 스트리밍 서비스는 없다. 오늘날 기기에서 신경망 비디오 디코더가 돌아간다면, 그것은 한 회사의 자체 앱 안에서다.

폐쇄 루프

인코더와 디코더가 모두 학습되는 엔드투엔드 신경망 코덱은 폐쇄 루프에서만 상용화된다. 한 회사가 양 끝을 통제하고 표준 코덱으로 폴백할 수 있는 환경이다. 마이크로소프트는 2021년부터 Teams와 Skype의 모든 일대일 통화에 AI 오디오 코덱을 써 왔고, 2026년 7월에는 현재 Teams P2P 통화에 확대 적용 중인 신경망 비디오 코덱 MLVC를 오픈소스로 공개했다. MLVC는 360p에서 같은 평가 품질을 내는 데 하드웨어 HEVC보다 비트레이트가 약 75% 적게 든다. 모든 음성·비디오 생성 모델은 학습된 코덱 위에서 돌아간다. 12.5 Hz 음성 토크나이저부터 압축비 1:192의 비디오 VAE까지 모두 그렇다. 신경망 텍스처 압축은 AAA 게임 한 편에 탑재되었다. 이들은 실제 배포 사례이며, 공통 조건이 하나 있다. 그 스트림을 다른 누구도 디코딩할 필요가 없다는 것이다.

시기

신경망 비트스트림의 대중적 재생은 표준과 실리콘 사이클을 기다린다. 신경망 도구를 실을 가능성이 큰 첫 표준, 즉 필자가 H.267이라 부를 VVC의 후속 표준은 2026년 7월 제안 요청을 발표했고, 2027년 1월 응답을 평가하며, 2029년 말 첫 버전을 목표로 한다. AV1은 규격 완성에서 넷플릭스 시청의 30%에 이르기까지 8년 가까이 걸렸다. 같은 패턴이라면 첫 H.267 칩은 2031년경 나오고, 대중화는 2030년대 중반에 이뤄진다 추정. 그사이 다음으로 화면에 도달할 코덱은 AV2인데, 2026년 5월 신경망 디코더 없이 완성되었고 IP 업체 세 곳이 3주 만에 AV2 디코더 고객을 발표했다.

실리콘

이유는 실리콘에 있다. 유일하게 공개된 신경망 비디오 디코더 칩 설계는 픽셀당 14.6 nJ의 에너지가 필요하다. DRAM까지 포함해 실측한 HEVC 디코더 칩의 약 12배, 코어만 놓고 보면 약 47배다. 마이크로소프트의 실시간 연구용 코덱 DCVC-RT는 논문의 모듈별 표를 필자가 합산해 보면 디코딩 픽셀당 약 17만 5천 회의 곱셈-누산(MAC)이 필요하며, 1080p30에서는 지속 연산 성능으로 약 22 TOPS에 해당한다 추정. AI PC를 정의하는 40 TOPS의 절반가량을 비디오 스트림 하나에 쓰는 셈이다. 신경망 디코더는 또한 모든 기기에서 비트 단위로 동일한 결과를 내야 하며, 그렇지 않으면 엔트로피 디코딩이 무너진다. MLVC가 NPU 간에 동일하게 디코딩하려고 압축 효율을 내주고, 4K가 아닌 540p로 동작하는 이유가 여기에 있다.

비즈니스

비즈니스 관점에서 돈은 독립형 신경망 코덱이 아니라 표준과 플랫폼을 따라간다. 플랫폼은 신경망 전처리·후처리에서 나오는 절감분을 가져간다. IP 업체는 새 표준마다 고정 기능 디코더를 라이선스하고, 여기에 이미징용 NPU를 더한다. 특허 풀은 기기당 로열티를 걷는다. VVC는 미국 등 Region 1 시장에서 휴대폰당 0.50달러, 4K TV당 1.00–1.50달러다. 가장 잘 알려진 신경망 코덱 스타트업 세 곳, WaveOne, iSIZE, Deep Render는 각각 애플, 소니, 인터디지털에 인수됐다. 열려 있는 기회는 신경망 전처리·후처리, 폐쇄 루프 제품, 고정 기능 코덱과 NPU를 짝지은 실리콘, 그리고 그 사이의 평가·표준화 작업에 있다.

질문	판단	핵심 수치(출처는 각 장에)
① 오늘날 신경망 코딩은 어디서 대규모로 상용화되었는가?	코덱 주변: 카메라 ISP, 인코딩 측 다운스케일링, 재생 시 초해상도. 엔드투엔드는 폐쇄 루프에서만: 통화, 게임 에셋, 생성형 AI 토크나이저	모바일 데이터 중 비디오 ≈75%, 2025년 출하된 휴대폰 12억 6천만 대와 TV ≈2억 대가 모두 고정 기능 하드웨어로 디코딩, DLSS 4는 250개 이상의 게임·앱에 적용
② 재생에서 신경망 코덱이 AV1, HEVC, VVC를 대체할 것인가? 언제인가?	표준과 기기 교체 주기 이전에는 아니다. H.267이 현실적인 첫 매개체이며, 전통 방식 도구와 신경망 도구를 섞을 가능성이 크다	H.267 제안 요청 2026년 7월, 평가 2027년 1월, 버전 1 2029년 말, 대중화 2030년대 중반 추정
③ 실리콘 비용은 얼마인가?	현재의 고정 기능 디코딩보다 픽셀당 에너지가 10배 수준 더 들고, 메모리 트래픽과 비트 단위 일치 요건이 더해진다	14.6 nJ/픽셀(신경망) 대 1.19(HEVC, DRAM 포함), 디코딩 픽셀당 ≈175 kMAC ≈ 1080p30에서 22 TOPS 추정
④ 표준과 특허는 무엇을 허용하는가?	신경망 후처리 필터는 이미 VVC 스트림에서 지정할 수 있다. JPEG AI는 ISO 표준이다. 어떤 신경망 도구가 라이선스 대상이 될지는 H.267이 결정한다	ECM(전통 방식)은 VVC 대비 ≈28%, 신경망 루프 필터는 10–22+ kMAC/픽셀에서 6–14%, 바이트댄스는 VVC 이후 세대 작업에서 특허 패밀리 645건
⑤ 누가 돈을 벌고, 기업은 어디서 기회를 찾을 수 있는가?	플랫폼, IP 업체, 특허권자. 순수 신경망 코덱 스타트업은 인수됐다. 기회는 전처리·후처리, 폐쇄 루프, 코덱+NPU 실리콘, 평가에 있다	VVC 풀 휴대폰당 0.50달러, 규격 발표 몇 주 만에 AV2 디코더 IP 라이선스, 인수 3건(조건 비공개)
⑥ 어떤 주장을 에누리해 들어야 하는가?	구형 코덱 대비, 낮은 해상도에서, 디코딩 비용 없이, 또는 출시 제품이 아닌 시연에서 나온 이득	MLVC: H.264 대비 88% 감소지만 하드웨어 HEVC 대비로는 65–76% 감소, 텍스처: 시연에서 VRAM 85% 절감, 출시 게임에서는 약 30%

그림 01

2026년 신경망 코딩이 상용화된 곳

신경망은 이미 주류 코덱의 양 끝에서 작동한다. 엔드투엔드 신경망 코덱은 한 회사가 인코더와 디코더를 모두 통제하는 곳에서만 상용화된다.

	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	—

출처: 보고서에 인용한 기업·표준·개발자 문서. 2026년 11월 12일 기준. 행은 각 영역이 닿는 사람 수 순으로 배열했다.

01 범위와 방법 ‘신경망 코딩’이 여기서 뜻하는 것과 살펴본 방법

신경망이 코덱에 들어가는 세 가지 방식

‘신경망 코딩’은 서로 다른 세 가지를 가리키며, 이를 둘러싼 논쟁 대부분은 이 셋을 뒤섞는 데서 생긴다.

01 **코덱 주변.** 신경망이 인코딩 전이나 디코딩 후에 픽셀을 처리하고, 비트스트림은 표준 그대로 유지된다. 예: 카메라의 AI 이미지 신호 처리기(ISP), 넷플릭스가 인코더 앞단에 둔 신경망 다운스케일러, TV의 초해상도. 어떤 기기든 그 스트림을 재생할 수 있으며, 신경망은 들어가는 것이나 나오는 것만 바꾼다.

02 **표준 내부.** 신경망이 표준 코덱의 도구가 된다. 예를 들어 모든 디코더가 실행해야 하는 신경망 인루프 필터나 인트라 예측기가 그렇다. 따라서 보통 정수 연산으로 정확하게 규정되어야 한다. 더 가벼운 형태로는 비트스트림이 메타데이터로 기술하고 디코더가 적용 여부를 선택할 수 있는 신경망 후처리 필터가 있다.

03 **코덱 대체.** 인코더와 디코더가 모두 엔드투엔드로 학습된다. 비트스트림은 짝이 맞는 신경망만 디코딩할 수 있는 양자화된 잠재값의 집합이다. 예: Teams의 마이크로소프트 MLVC, 음성 모델의 음성 토크나이저, 게임의 신경망 텍스처.

그림 02 **신경망이 코덱에 들어가는 세 가지 방식**

2026년에 상용화된 것은 대부분 표준 코덱 주변에서 작동한다. 표준 내부의 신경망 도구는 규격화가 진행 중이다. 엔드투엔드 신경망 코덱이 코덱을 대체하는 곳은 한 회사가 양 끝을 모두 통제하는 경우뿐이다.



출처: 이 보고서의 출처를 저자가 종합. 2026년 11월 12일 기준.

첫 번째 방식은 누구의 허락도 필요 없다. 두 번째는 표준이 필요하다. 세 번째는 양 끝을 모두 통제해야 한다. 이것이 사람들에게 닿는 순서이며, 이후 내용의 대부분을 설명해 준다.

연구 의제가 아니라 현실 세계의 순서대로

연구 논문은 주로 세 번째 방식을 다룬다. 압축 기록이 나오는 곳이 거기이기 때문이다. 이 보고서는 각 영역이 닿는 사람 수에 따라 순서를 정했다. 돈과 엔지니어링 역량이 어디로 갈지를 결정하는 것이 바로 그것이기 때문이다. 재생과 스트리밍이 먼저이고, 이어서 사진, 통화, 생성형 AI, 게임 순이며, 특수 링크(위성, 드론, 로봇)가 마지막이다. 에릭슨 집계로 2025년 말 비디오는 전체 모바일 데이터 트래픽의 약 75%를 차지한다. 2025년 업계는 스마트폰 약 12억 6천만 대(IDC)와 TV 약 1억 9,600만–2억 900만 대(트렌드포스와 옴디아 추정)를 출하했다. 이 기기들은 모두 고정 기능 하드웨어로 비디오를 디코딩한다.

조사 방법

이 보고서는 칩이나 제품 계획을 검토하는 방식으로 진행한 문헌 조사다. 표준 문서와 회의 보고서(JVET, MPEG, JPEG, AOMedia, IETF), 기업의 엔지니어링 블로그와 릴리스 노트, 칩 논문, 특허 풀 요율표, 수출 통제 문서, 언론 보도를 읽었다. 모든 수치에는 날짜를 붙였다. 논문의 표에서 픽셀당 연산량을 계산한 것처럼 필자가 직접 도출한 수치는 그렇다고 밝히고 추정으로 표시했다. 1차 출처로 확인할 수 없었던 주장은 업체 주장으로 표시했다. 금액은 미국 달러 기준이다. 현황은 2026년 11월 12일 기준이다.

가능한 곳에서는 같은 조건끼리 비교했고, 그럴 수 없는 곳에서는 그렇다고 밝혔다. 압축 이득은 네 가지가 함께 제시되지 않으면 의미가 거의 없다. 기준(무엇 대비 이득인가), 지표(PSNR, VMAF 또는 시청자 패널), 콘텐츠와 해상도, 그리고 디코딩 비용이다. 이 네 가지만 밝히면 신경망 코덱을 둘러싼 이견 대부분이 사라진다.

02장

02 재생과 스트리밍: 양 끝은 신경망, 가운데는 표준

재생과 스트리밍

시청자가 받는 것은 여전히 표준 비트스트림이다

최대 비디오 서비스들은 코덱 전환을 한 차례 마치고 다음 전환을 시작하고 있으며, 어느 쪽에도 신경망 디코더는 없다. 넷플릭스는 2025년 12월 AV1이 자사 시청량의 약 30%를 차지해 AVC에 이어 두 번째로 많이 쓰이는 코덱이 되었다고 발표했다. AV1은 하드웨어 디코더가 TV에 들어간 뒤에야 TV에 도달했다. 넷플릭스는 2019년 기기 인증 항목에 AV1을 추가했고, 2021년 말부터 TV로 AV1 스트리밍을 시작했으며, 2021년부터 2025년까지 인증을 신청한 대화면 기기의 88%가 AV1을 지원했는데 거의 모두 2023년 이후 기기였다고 밝혔다. 유튜브는 2026년 3월 19일 기준으로 시청 시간 가중 카탈로그의 90% 이상을 AV1 형식으로 제공한다고 밝혔다.

AV1의 후속 코덱으로 Alliance for Open Media가 내놓은 AV2는 2026년 5월 28일 버전 1.0.0에 도달했으며, AOMedia의 객관적 측정으로 AV1보다 압축 효율이 약 30% 높다. 유튜브는 웹에서 AV2 스트리밍을 “이르면 2027년 초”에 시작할 계획이다. 실리콘은 곧바로 뒤따랐다. 칩스앤미디어는 5월 AV2 하드웨어 디코더 개발을 마치고 6월 이름을 밝히지 않은 북미 빅테크 기업에 라이선스했다. Allegro DVT는 6월 16일까지 AV2 디코더 IP를 고객 세 곳에 공급했고, 베리실리콘은 6월 10일까지 AV2를 지원하는 비디오 IP를 여러 고객에게 출시했다. AV2는 전통 방식 코덱이다. 이득은 엔드투엔드 신경망이 아니라 더 나은 분할, 예측, 변환, 인루프 필터(그중 하나는 “데이터 기반” 가이드 디테일 필터)에서 나온다.

현행 ITU/ISO 표준인 VVC는 브라질에서 첫 방송 무대를 찾았다. 브라질의 DTV+(TV 3.0) 시스템은 월드컵에 맞춰 2026년 6월 상업 운영에 들어갔다. 2026년 7월 기준으로 브라질에서 판매 중인 TV 가운데 DTV+ 수신기를 내장한 제품은 없었으며, 첫 제품은 2026년 말에서 2027년 초 사이에 나올 것으로 예상된다.

인코더 쪽: 넷플릭스의 신경망 다운스케일러

재생 분야에서 가장 깔끔한 신경망 코딩 사례는 2022년부터 운영되어 왔으며, 시청자의 기기가 해야 할 일은 아무것도 바꾸지 않는다. 넷플릭스는 타이틀마다 여러 해상도로 인코딩하는데, 4K 마스터에서 1080p, 720p, 540p를 만드는 다운스케일링 단계에는 전통적으로 Lanczos 같은 고정 필터를 썼다. 넷플릭스는 이를 작은 신경망으로 바꿨다. 전처리 블록 뒤에 리사이징 블록이 오는 구조로, 기기의 표준 업스케일러를 거친 뒤에 보이는 화질을 높이도록 학습되었다. 넷플릭스는 이를 그대로 끼워 넣으면 되는 해법이라고 설명한다. 인코딩 쪽이나 클라이언트 쪽에서 다른 변경이 필요 없으므로 "넷플릭스 스트리밍을 지원하는 수백만 대의 기기가 자동으로 혜택을 본다"는 것이다. 넷플릭스의 선호도 시험에서 시청자의 약 77%가 신경망 다운스케일러의 결과물을 택했다 업체 주장. 넷플릭스 규모에서 비용을 감당할 수 있는 수준으로 유지하는 데는 "신경망 레이어 몇 개면 충분했다".

기억해 둘 패턴이 바로 이것이다. 신경망은 연산이 저렴하고 일정을 조정할 수 있는 데이터센터에서 타이틀당 한 번 실행되며, 그 출력은 평범한 AV1 또는 HEVC 스트림이다.

화면 쪽: 앱과 TV 안의 초해상도

반대쪽 끝에서는 신경망이 코덱과 원본이 잃은 것을 복원한다. 유튜브는 2025년 10월 29일 TV 앱용 Super Resolution을 발표했다. 1080p 미만(240p~720p)으로 업로드된 동영상을 머신러닝으로 HD까지 업스케일링하며, 4K를 목표로 내걸었다. 업스케일링된 버전에는 표시가 붙고, 시청자는 원본으로 되돌릴 수 있으며, 크리에이터는 적용을 거부할 수 있다. 메타는 2025년 10월, 업로드 시 서버에서(인텔과 협력해 표준 x86 CPU에서) 그리고 재생 시 클라이언트에서 비디오 초해상도를 모두 운영하는 방식을 설명하면서, 두 방식이 서로 보완한다는 점을 확인했다고 밝혔다. TV 제조사도 실리콘으로 같은 일을 한다. 삼성전자의 플래그십 8K 프로세서 NQ8 AI Gen3는 업스케일링용 신경망 768개를 내세운다.

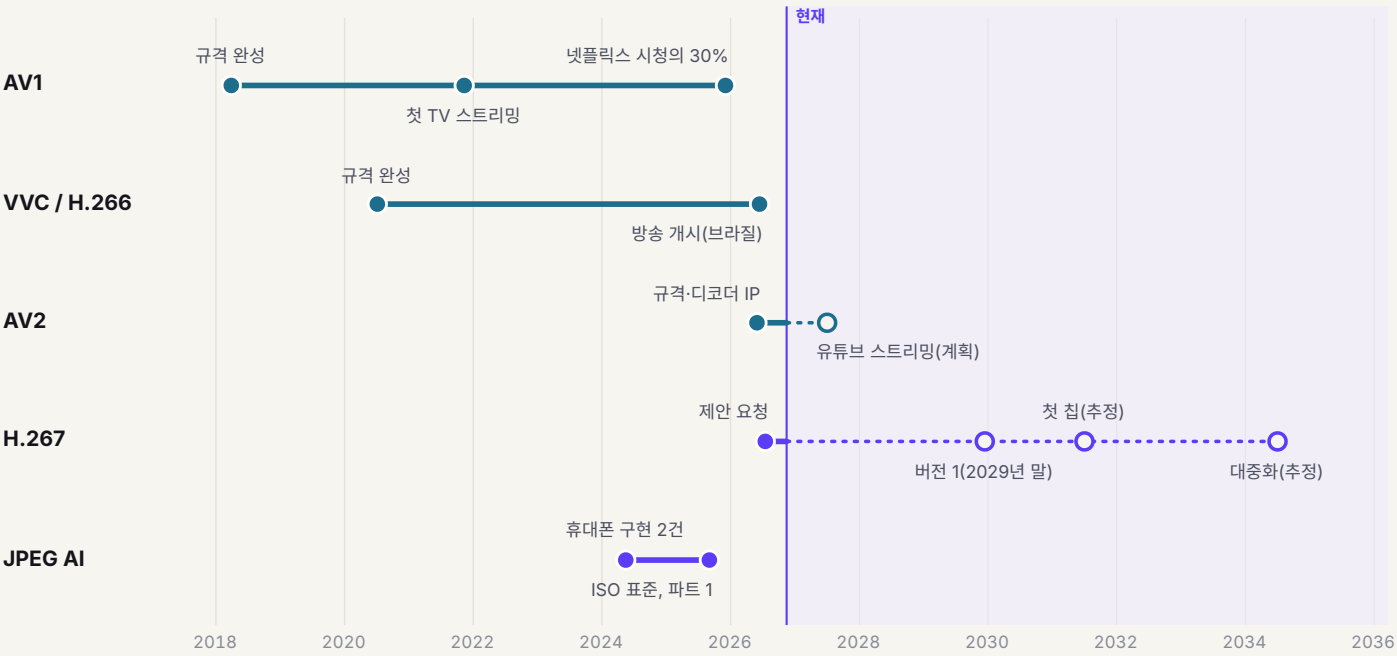
경제성은 신경망이 어디서 도는지에 따라 달라진다. 서버 측 처리는 플랫폼이 손대는 모든 동영상마다 연산 비용을 치르게 한다. 기기 측 처리는 플랫폼에 비용이 들지 않는다. 구매자가 이미 값을 치른 NPU를 쓰기 때문이다. 2026년의 가장 큰 이득이 양쪽 모두에서 나오는 이유다.

신경망 비트스트림을 디코딩하는 TV가 없는 이유

재생 포맷은 그것을 재생할 모든 기기에서 디코딩되어야 하고, 그 기기들은 수년간 쓰인다. 그래서 전통 방식 코덱조차 표준 완성에서 대중 기기 보급까지 가는 길이 느리다. AV1 규격은 2018년 3월 완성되었다. 몇 달 안에 dav1d 소프트웨어 디코더가 뒤따랐고, 2021년 말부터 TV로 스트리밍되었으며, 규격 완성 후 8년 가까이 지난 2025년 12월 넷플릭스 시청의 약 30%에 이르렀다. VVC는 2020년 7월 완성되어 2026년 6월 첫 방송 서비스에 도달했다.

그림 03 **표준 완성에서 대중 기기 보급까지 10년 가까이 걸린다**

재생에는 기기 안의 하드웨어 디코더가 필요하다. AV1은 규격 완성에서 넷플릭스 시청의 30%에 이르기까지 8년 가까이 걸렸다. 신경망 도구를 실을 가능성이 큰 첫 표준은 2029년 말에 나올 예정이다.



출처: Netflix Tech Blog(2025년 12월); AOMedia; Allegro DVT; TV Technology; Fraunhofer IIS(브라질 DTV+); MPEG-155; JPEG 위원회; ISO. 속이 빈 표시는 목표치 또는 저자의 추정이다.

신경망 비트스트림에는 전통 방식 코덱이 오래전에 해결한 두 가지 문제가 더해진다. 픽셀당 에너지와 비트 단위로 일치하는 디코딩이다. 둘 다 8장에서 다룬다. 이 두 문제 때문에, 무엇을 만들어야 하는지 표준이 정확히 알려 주기 전에는 어떤 TV나 휴대폰 제조사도 모든 기기에 신경망 비디오 디코더를 넣지 않을 것이다.

판을 바꿀 수 있는 것: H.267

신경망 도구를 모든 화면에 실어 나를 수 있는 표준은 ITU-T와 ISO/IEC가 JVET(Joint Video Experts Team)에서 공동 개발하는 VVC의 후속 표준이다. 이름은 아직 확정되지 않았다. 업계의 많은 이들처럼 필자도 이를 H.267이라 부르겠다.

2025년 10월 JVET은 증거 제출 요청에 대한 응답을 평가했다. 16개 기업과 연구소가 응답 5건을 냈고, 코덱 구성 20개를 시험했다. 위원회는 VVC보다 압축 효율이 높다는 증거를 확인했으며, 개선된 기존 기술과 신경망 기술을 모두 "흥미롭다"고 평가했다. 신경망 방식은 특히 8K HDR 콘텐츠에서 뚜렷한 지각 화질 이득을 보였고, 게임 콘텐츠에서는 이득이 더 작았다. 응답자 중 하나인 인터디지털은 기존 방식의 ECM(Enhanced Compression Model) 기반 설계 하나와 VVC 위에 AI 도구를 얹은 설계 하나를 제출했다.

공식 제안 요청은 2026년 7월 제네바 회의에서 발표되었다. 주 제안서 마감은 2026년 10월 말이었고, 보충 자료는 12월 중순이 마감이며, 응답은 2027년 1월 브리즈번 JVET 회의에서 평가된다. JVET은 2027년 하반기에 첫 테스트 모델을, 2029년 말까지 표준의 첫 버전을 내놓을 것으로 예상하며, 이 작업을 "새 표준을 시작하기 위해 이제껏 기울인 것 중 아마도 가장 큰 노력"이라고 부른다.

전통 방식 트랙이 세운 기준은 높다. 기존 신호 처리로 만든 ECM은 랜덤 액세스 구성에서 VVC 대비 약 28%를 절감한다. JVET의 신경망 탐색 연구는 픽셀당 1만 회의 곱셈-누산으로 약 6%, 2만 2천 회로 8%, 고복잡도 동작점에서 14%를 절감하는 인루프 필터를 보고했다. 필자의 판단으로는 H.267은 절감되는 비트가 디코더 비용을 정당화하는 곳에만 신경망 도구 몇 개를 넣은 전통 방식 코어가 될 가능성이 가장 크며, 그 수가 몇 개일지는 2027년 1월 평가에서 드러날 것이다 분석.

이미 가교는 놓여 있다. VVC와 HEVC 스트림은 디코더 측에서 적용할 신경망 후처리 필터를 기술하는 부가 향상 정보(SEI)를 실을 수 있으며, 2025년 10월 완성된 VVC와 그 SEI 표준(VSEI)의 새 판은 이 메시지를 확장했다. 스트림은 모든 디코더가 실행하도록 강

제하지 않으면서 디코딩 후에 어떤 신경망을 실행할지 알려 줄 수 있다.

재생 분야 비즈니스에 주는 의미

지금은 코덱 주변에서 작동하는 것을 팔아라. 서비스 업체에는 인코딩 측 전처리를, TV와 셋톱 칩에는 후처리 블록과 튜닝을 팔면 된다. 2030년대 이전에 타사 기기에서 디코딩되어야 하는 신경망 비트스트림 위에 재생 사업을 세우지 마라. 실리콘을 만든다면 단기 매출은 AV2와 VVC 디코더에서 나온다. 신경망 부분은 그 옆에 두는 화질 개선용 NPU 경로다(8장).

03장

03 사진과 이미지: 카메라 안은 신경망, 웹은 전통 방식

사진과 이미지

카메라 파이프라인은 이미 신경망이다

최근 플래그십 휴대폰으로 찍은 사진은 모두 압축되기 전에 신경망을 거친다. 2023년 휴대폰부터 탑재된 퀄컴 Snapdragon 8 Gen 2는 사진·동영상 촬영 중에 실시간 시맨틱 세그멘테이션을 수행해 얼굴, 머리카락, 하늘, 배경을 따로 조정하는 “Cognitive ISP”를 도입했다. 삼성전자 Galaxy S23 Ultra는 이를 셀피 카메라에 적용한 첫 휴대폰이었다. 2026년 1월 칩스앤미디어와 Visionary.ai는 CES에서 스스로 최초의 완전 AI 기반 ISP라 부르는 제품을 선보였다. 고정 기능 ISP 없이 칩스앤미디어의 NPU인 WAVE-N에서만 16비트 부동소수점으로 동작하고 OTA로 업데이트할 수 있는 엔드투엔드 이미징 파이프라인이다. Visionary.ai의 CEO는 “기존 파이프라인은 고정 기능 ISP에 신경망 블록을 덧붙인다. 우리는 기존 ISP를 완전히 대체한다”고 말했다. 이 ISP는 전시회가 끝난 뒤 제조사들에 평가용으로 제공되었다.

그런 다음 결과물은 어떤 기기든 열 수 있는 전통 방식 포맷인 JPEG이나 HEIF로 저장된다. 패턴은 재생과 같다. 한 회사의 파이프라인 안에서는 신경망, 경계에서는 표준이다.

JPEG AI: 완성되었지만 아직 제품이 없는 신경망 표준

JPEG AI는 ISO/IEC와 ITU-T의 첫 학습 기반 이미지 코딩 표준이다(ISO/IEC 6048, ITU-T T.840). 핵심 코딩 시스템인 파트 1은 2025년에 발행되었다. JPEG 위원회는 픽셀당 7천에서 21만 6천 회의 곱셈-누산에 이르는 디코더 복잡도 범위에서 VVC 인트라 코딩 대비 12.5%에서 27.9%의 압축 이득을 보고했다. 이 설계는 공동으로 학습된 인코더 2개와 디코더 3개로 구성되어, 수신기가 디코더를 하나만 구현해도 모든 스트림을 읽을 수 있다. 위원회는 2024년 5월까지 스마트폰 구현 2건을 보고했다. 이와 별도로 IEEE는 2024년 12월 NPU용으로 설계된 신경망 기반 이미지 코딩 표준 IEEE 1857.11-2024를 발행했다.

2026년 11월 기준으로 JPEG AI를 기본으로 쓰거나 읽는 휴대폰 카메라, 사진 앱, 브라우저는 찾지 못했다.

웹이 대신 채택한 것: JPEG XL

2026년에 실제로 입지를 넓힌 이미지 포맷은 전통 방식이다. 구글은 실험적 지원을 제거한 지 4년 만인 2026년 10월 6일, Rust로 작성한 새 메모리 안전 디코더를 써서 Chrome 155에서 JPEG XL 디코딩을 기본 활성화했다. Safari는 2023년부터 JPEG XL을 디코딩해 왔고, 모질라도 같은 방향으로 움직여 왔다. 구글이 내세우는 근거는 JPEG보다 나은 압축, 기존 JPEG 파일의 무손실 재압축, HDR, 점진적 디코딩이다.

이 교훈은 일반화할 수 있다. 어디서나 열려야 하는 포맷은 어떤 CPU에서도 돌아가고 메모리 안전하게 만들 수 있는 빠르고 결정론적인 디코더로 이긴다. 신경망 이미지 디코더에는 NPU나 GPU가 필요하고, 기기 간에 정확히 일치하는 연산이 필요하다. 신경망 코

딩은 기기 자체의 파이프라인 안에서, 그리고 같은 회사가 인코딩과 디코딩을 모두 하는 시스템에서 이긴다. 압축 파일을 다른 누구에게도 넘기지 않는 클라우드 사진 서비스가 그런 예다.

이미징 분야 비즈니스에 주는 의미

돈은 촬영 파이프라인에 있다. AI ISP IP, 이미징에 맞춘 NPU, 그리고 카메라 제조사가 사는 튜닝 작업이다. 저장과 전송에서 학습 기반 이미지 코덱은 현재로서는 폐쇄형 시스템용 도구일 뿐, 승부를 걸 포맷이 아니다.

04장

04

통화와 회의

통화와 회의: 엔드투엔드 신경망 코덱이 실제로 상용화된 곳

오디오가 먼저였다

마이크로소프트는 2021년 2월 AI 기반 음성 코덱 Satin을 소개하면서, 이미 Teams와 Skype의 모든 일대일 통화에 쓰이고 있다고 밝혔다. Satin은 6 kbps에서 최대 16 kHz 오디오 대역폭의 초광대역 음성을 전달한다. Skype의 이전 코덱 Silk는 같은 비트레이트에서 협대역이었다. 절감한 비트레이트는 중복 전송에 쓰여 통화가 패킷 손실을 더 잘 견딘다.

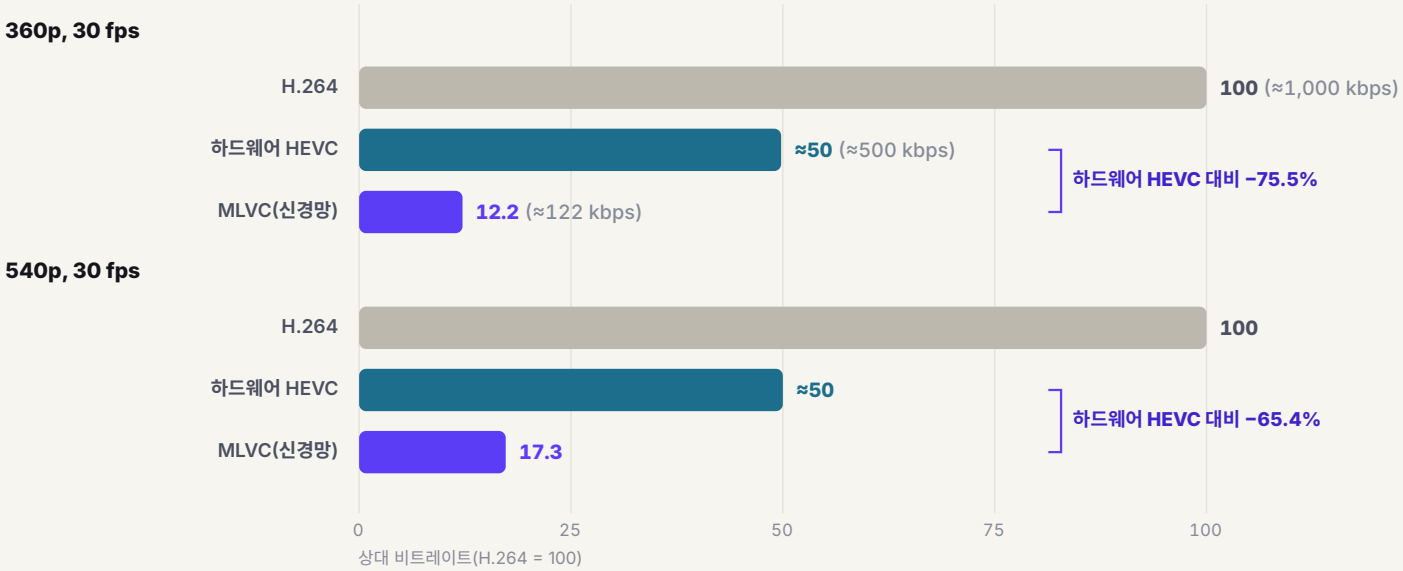
MLVC: 실서비스 통화에 쓰이는 신경망 비디오 코덱

2026년 7월 24일 마이크로소프트는 자사의 머신러닝 비디오 코덱 MLVC를 MIT 라이선스의 오픈소스로 공개했다. 모델 코드, 학습된 가중치, 학습 스크립트, 여러 NPU를 대상으로 하는 변환 도구가 함께 공개되었다. MLVC는 마이크로소프트 리서치가 2021년부터 발표해 온 DCVC 계열의 제품 버전이다. 마이크로소프트는 실시간 원격 측정, A/B 테스트, 그리고 폴백용 기존 코덱을 갖추고 Teams의 P2P 통화에 MLVC를 확대 적용하고 있다고 밝혔다.

MLVC는 540p와 360p에서 초당 30프레임으로 동작하며, 애플 Neural Engine과 퀄컴·인텔 NPU에서 NPU를 절반 이하로만 쓰도록 CPU와 NPU에 작업을 나눠 설계되었다. 마이크로소프트 자체 Video Conferencing Dataset에서 ITU-T P.910에 따라 실시한 주관 평가에서, MLVC는 360p에서 H.264보다 87.8%, 하드웨어 HEVC보다 75.5% 적은 비트레이트가 필요했고, 540p에서는 각각 82.7%와 65.4% 적었다 업체 주장. 360p에서 이는 H.264가 약 1 Mbps를 쓰는 자리에 약 122 kbps를 쓰는 셈이다.

그림 04 **통화에서 신경망 코덱은 하드웨어 HEVC 대비 비트레이트를 65–76% 줄인다**

마이크로소프트의 화상회의 테스트셋에서 같은 평가 품질에 필요한 비트레이트(H.264 = 100). MLVC는 노트북과 휴대폰의 NPU에서 최대 540p30으로 동작하며 NPU를 절반 이하로만 쓴다.



출처: 마이크로소프트, MLVC 공개 자료(2026년 7월 24일): 자사 Video Conferencing Dataset에 대한 ITU-T P.910 주관 평가. HEVC 막대는 발표된 절감률에서 역산했다. 독립 측정이 아닌 업체 주장이다.

헤드라인 수치는 두 가지를 감안해 읽어야 한다. 88%는 H.264 대비다. 같은 노트북에 이미 들어 있는 하드웨어 HEVC 대비로는 65–76%로, 여전히 크다. 그리고 MLVC는 배포할 수 있도록 압축 효율을 포기했다. 한 독립 분석에 따르면 지각 품질 기준으로 학습한 DCVC-RT는 같은 척도에서 81.9%를 기록하는데, MLVC는 75.5%다. 마이크로소프트가 소유하지 않은 하드웨어에서 똑같이 디코딩하는 대가는 약 6%포인트였다.

통화가 자연스러운 무대인 이유

통화는 폐쇄 루프다. 양쪽 단말은 같은 애플리케이션을 실행하고, 통화 시작 시 서로 무엇을 할 수 있는지 협상하며, 통화 도중 표준 코덱으로 폴백할 수 있다. 해상도는 높지 않고, 네트워크는 열악한 경우가 많으며, 휴대폰이 몇 년 전부터 그랬듯 이제는 양쪽 노트북에도 NPU가 탑재되어 출시된다. 마이크로소프트는 AI PC의 기준으로 40 TOPS NPU를 제시하고, 애플 M4의 Neural Engine은 38 TOPS로 평가된다. 재생에서 신경망 코덱을 가로막는 조건이 여기에는 하나도 없다.

반례: Opus와 MLow

모든 통화 플랫폼이 같은 선택을 하지는 않았고, 그 이유가 시사적이다.

WebRTC와 대부분의 통화 앱에 쓰이는 개방형 오디오 코덱 Opus는 버전 1.5(2024년 3월)에서 딥러닝을 추가했다. 신경망으로 신호 자체를 처리하거나 생성한 것은 이때가 처음이다. 하지만 선택 기능으로 추가했다. 비트스트림 변경이 필요 없는 디코더 측 음질 개선(LACE와 NoLACE), 신경망 패킷 손실 은닉, 딥 리던던시 방식(DRED)이다. 모두 기본값은 꺼져 있어 오래된 기기에 부담을 주지 않는다. Opus 1.6은 신경망 대역폭 확장과 크기를 3분의 1로 줄인 DRED 모델을 추가했으며, 2012년 표준과 완전한 호환성을 유지한다. 이 접근법은 '표준 내부, 선택 사항, 하위 호환'이다.

메타는 한 걸음 더 물러섰다. 2024년 6월, 신경망이 아닌 전통적인 신호 처리로 만든 저비트레이트 음성 코덱 MLow를 출시한 것이다. 현재 Instagram과 Messenger의 모든 통화에 쓰이며 WhatsApp에도 확대 적용 중이다. 메타가 든 이유는 사용자의 기기였다. 메타 통화의 20% 이상이 ARMv7 휴대폰에서 이뤄지고, 하루 수천만 건의 WhatsApp 통화가 10년 넘은 휴대폰에서 이뤄진다. 메타

는 자사의 EnCodec을 포함한 신경망 코덱이 비싼 단말에서만 안정적으로 동작한다고 썼다. MLow는 6 kbps에서 Opus의 약 2배 품질을 내면서 연산량은 10% 적다.

사용자가 플래그십 노트북부터 10년 된 휴대폰까지 걸쳐 있는 서비스에서는 복잡도 예산이 결정을 내린다. 신경망 코덱은 보급된 기기가 새것이고 업체가 양 끝을 통제하는 곳에서 먼저 상용화된다.

통신 분야 비즈니스에 주는 의미

기업용 화상회의, 텔레프레즌스, 원격 지원, 중계 링크, 사설망은 양 끝이 모두 자사 것이라면 엔드투엔드 신경망 코덱이 지금 바로 제품이 될 수 있는 곳이다. 처음부터 세 가지를 계획하라. 폴백 코덱, 지원하는 모든 NPU에서의 디코딩 시험, 그리고 업체 데이터셋이 아닌 자사 콘텐츠로 하는 품질 평가다.

05장

05

생성형 AI

생성형 AI: 아무도 코덱이라 부르지 않는 최대 규모의 배포

모든 음성·비디오 모델은 학습된 코덱 위에서 돌아간다

엔드투엔드 신경망 압축의 최대 배포처는 어떤 코덱 표준에도 없다. 생성형 AI 안에 있다. 음성 모델은 파형을 직접 듣거나 만들어 내지 않는다. 학습된 오디오 코덱의 이산 토큰을 읽고 쓴다. 2024년 공개된 Kyutai의 Mimi는 음성을 초당 12.5프레임, 1.1 kbps의 토큰으로 바꾼다. 어떤 전화 코덱보다 낮은 비트레이트이며, 의미 정보와 음향 정보를 하나의 스트림에 담는다. 2022년 발표된 메타의 EnCodec도 그런 예다. 메타의 엔지니어들은 나중에 이런 코덱이 낮은 비트레이트에서는 뛰어나지만 오래된 휴대폰에는 너무 무겁다고 평했다(4장).

비디오 생성도 같은 방식이다. 확산 모델이나 트랜스포머 모델은 학습된 비디오 오토인코더(VAE)의 잠재 공간에서 생성하고, VAE의 디코더가 그 결과를 픽셀로 바꾼다. LTX-Video의 VAE는 비디오를 1:192로 압축한다. 잠재 토큰 하나가 128채널로 8프레임에 걸친 32×32 픽셀 블록을 나타낸다. 이 디코더는 마지막 노이즈 제거 단계도 수행하므로, 디테일을 복원만 하는 것이 아니라 일부는 생성한다.

이들이 배포용 포맷이 아닌 이유

이 코덱들은 음성·비디오 모델에 요청이 들어올 때마다 실행되지만, 다른 누군가가 디코딩하는 포맷은 하나도 없다. 모델마다 고유한 잠재 공간이 있다. 레이트 제어도, 적합성 시험도, 다른 디코더가 같은 영상을 만든다는 보장도 없다. 주어진 비트레이트에서의 충실도가 아니라 생성 품질을 위해 설계되었고, 데이터센터 GPU에서 디코딩된다. 이들의 성공은 한 회사가 인코더와 디코더, 하드웨어를 모두 소유하면 학습 기반 압축이 대규모로 작동한다는 것을 보여 준다.

비즈니스에 주는 의미

실무적인 시사점은 두 가지다. 첫째, 생성형 비디오나 음성을 서비스하는 쪽에게 토크나이저 효율은 곧바로 비용 항목이다. 모든 토큰에 걸쳐 어텐션을 수행하는 트랜스포머에서는 연산량이 토큰 수보다 빠르게 늘어나므로, 압축률이 높은 잠재 표현일수록 생성 비용이 비례 이상으로 줄어든다 분석. 잠재 표현 설계, 엔트로피 모델, 디코더 효율을 이해하는 코덱 엔지니어는 생성형 AI 팀에 귀중한 인재다. 둘째, 경계가 움직이고 있다. 표준화 기구들은 사람이 아닌 기계를 위한 코딩 작업에 착수했다. MPEG의 Video Coding for Machines(ISO/IEC 23888-2)는 2026년 초 국제 표준안 단계에 도달했고 2026년 말 최종 확정될 예정이며, JPEG AI는 처음부터 사람의 시청뿐 아니라 컴퓨터 비전 작업에도 쓰이도록 설계되었다.

06장

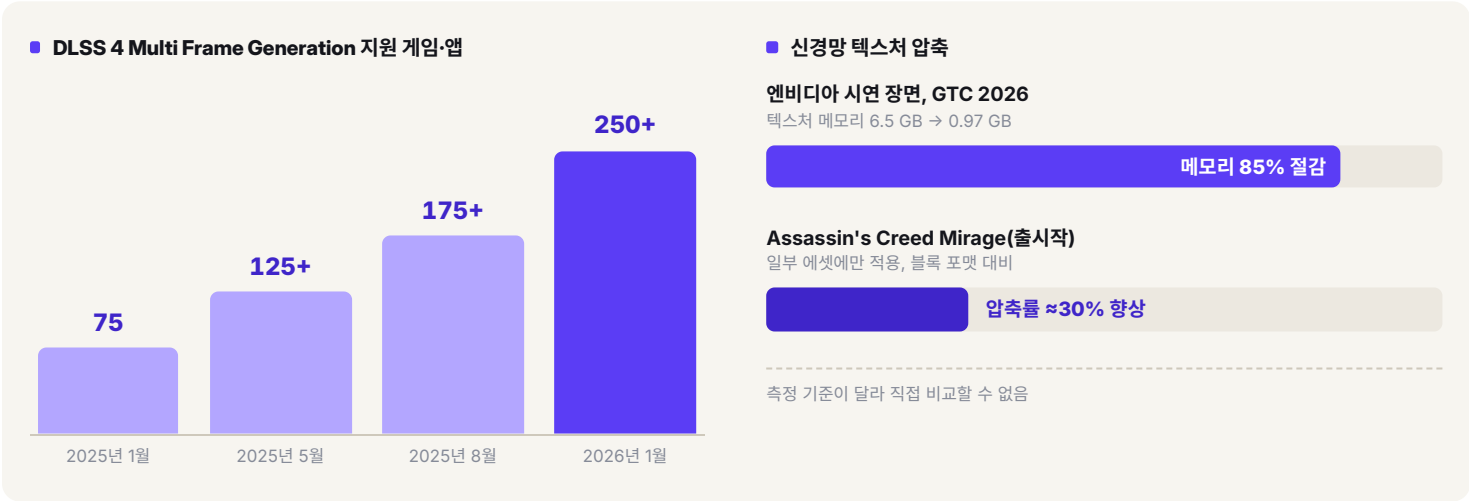
06 게임과 그래픽: 신경망 복원은 주류, 신경망 에셋은 시작 단계

픽셀을 덜 렌더링하고 나머지는 신경망이 채운다

실시간 그래픽에서는 이미 신경망이 플레이어가 보는 화면의 대부분을 재구성한다. 엔비디아의 DLSS는 픽셀과 프레임을 덜 렌더링하고 나머지는 신경망으로 만들어 낸다. Multi Frame Generation을 갖춘 DLSS 4는 2025년 1월 게임·앱 75개의 지원과 함께 출시되었다. 엔비디아 집계로 지원 수는 2025년 5월 125개 이상, 8월 175개 이상으로 늘었고, 2026년 1월에는 250개 이상이 되었다. 엔비디아는 이때 초해상도용 2세대 트랜스포머 모델을 갖춘 DLSS 4.5를 발표했다. 이름만 다를 뿐 사실상 디코더다. '비트스트림'은 저해상도 프레임과 모션 데이터이고, 신경망이 전체 이미지를 재구성한다.

그림 05 **한 회사가 스택 전체를 통제하면 신경망 복원은 빠르게 확산된다**

왼쪽: 엔비디아 DLSS 4 Multi Frame Generation을 지원하는 게임과 앱. 오른쪽: 시연 장면과 출시 게임의 신경망 텍스처 압축. 출시 게임은 런타임 비용 때문에 일부 에셋에만 적용했다.



출처: 엔비디아(2025년 1월부터 2026년 1월까지, GTC 2026 시연); Ubisoft La Forge(2026년 2월). 업체 주장이며, 두 텍스처 수치는 서로 다른 것을 측정했다.

신경망 텍스처는 선별적으로 탑재되었다

텍스처는 GPU가 하드웨어로 디코딩하는 고정 블록 포맷(BCn)으로 압축된다. 신경망 텍스처 압축은 재질 텍스처 묶음을 간결한 잠재 표현과, 필요할 때 텍셀을 재구성하는 작은 신경망으로 대체한다. 2026년 2월 유비소프트의 La Forge 연구소는 2023년 게임 Assassin's Creed Mirage에 이를 탑재한 경험을 소개하며, 같은 시각 품질에서 기존 포맷보다 실효 압축률이 약 30% 높았다고 밝혔다 업체 주장. 모든 곳에 쓰지는 않았다. "사용은 런타임 성능의 제약을 받았기" 때문에, 유비소프트는 인스턴스 수가 많고 메모리 압박이 큰 일부 에셋에만 적용했다.

GitHub에서 베타로 제공되는 엔비디아의 RTX Neural Texture Compression SDK는 상한을 보여 준다. GTC 2026 시연에서 BCn으로 6.5 GB의 텍스처 메모리를 쓰던 빌라 장면은 신경망 텍스처로 970 MB만 필요했다. 85% 줄어든 것이다 업체 주장. 이 SDK에는 두 가지 경로가 있다. 로드 시점에 BCn으로 디코딩하면 디스크 공간과 다운로드 크기는 줄지만 비디오 메모리는 줄지 않는다. 샘플링 시점에 디코딩하면 메모리는 줄지만 매 프레임 셰이더 시간이 든다.

API가 도착하고 있다

그래픽 API는 셰이더 안에서 신경망 디코딩에 필요한 행렬 연산을 추가하고 있다. 마이크로소프트는 2026년 2월 Shader Model 6.9를 출시했다. 2025년 프리뷰로 선보인 cooperative vector 기능은 이후 셰이더 모델을 위한 통합 선형대수 설계에 흡수되었고, DirectX Linear Algebra가 2026년 4월 퍼블릭 프리뷰에 들어갔다. 이것이 모든 드라이버에 들어가면 신경망 텍스처 디코더는 하나의 코드 경로로 어느 업체의 GPU에서든 실행될 수 있다.

비즈니스에 주는 의미

게임이 신경망 디코딩을 탑재하는 것은 폐쇄 루프의 조건이 성립하기 때문이다. 엔진 하나, 에셋 파이프라인 하나, GPU의 텐서 하드웨어, 그리고 폴백 포맷이 있다. 남은 과제는 도구, 신경망 재질을 인코딩하는 에셋 파이프라인, 업체 간 호환 런타임, 그리고 무엇보다 모바일이다. 모바일에서는 디코딩된 텍셀당 에너지 비용이 승부를 가른다.

07장

07 특수 링크: 위성, 드론, 로봇

이론상 엔드투엔드 신경망 코딩이 가장 잘 맞는 경우들이다. 한 소유자가 링크의 양 끝을 통제하고, 대역폭은 부족하고 손실이 많으며, 콘텐츠 범위가 좁아 학습시키기 쉽고, 지각적으로, 나아가 의미적으로만 충실한 영상으로도 충분한 경우가 많다. 이런 링크를 위한 손실에 강한 신경망 코덱은 활발한 연구 분야이며, 저비트레이트 영상의 서버 측 복원은 실제로 쓰이고 있다.

동시에 소비자 비디오에 비하면 작고 규제가 점점 강해지는 시장이다 분석. 중국에서는 상무부와 다른 두 기관이 10월 19일 발표하고 2026년 11월 18일부터 시행하는 2026년 제49호 공고가 드론 관련 품목 13개 범주의 수출 통제를 개정한다. 업계 매체 보도에 따르면, 기술 항목 가운데 하나는 시맨틱 이미지 압축률 75%, 150밀리초 이내 이미지 복원 같은 기준치로 일부 정의되는 무인 지상 플랫폼을 다룬다. 시맨틱 압축 자체가 수출 통제 파라미터가 된 것이다. 미국에서는 연방통신위원회(FCC)가 2025년 12월 22일 외국산 드론과 그 핵심 부품을 Covered List에 올렸고, 이에 따라 새로운 외국산 기종은 더 이상 장비 인증을 받을 수 없다.

결론: 산업·정부 고객을 두고 양쪽의 수출 규정 준수를 관리할 역량을 갖춘 전문 기업에는 탄탄한 틈새시장이지만, 주류 시장은 아니며, 신경망 코딩이 대규모로 작동하는지 알아볼 곳도 아니다.

08장

08 실리콘 관점: 에너지, 연산, 메모리, 정확성

필자의 이력이 가장 중요해지는 장이다. 필자는 풀 커스텀 ASIC 데이터패스 설계, 비디오 코덱 가속기 같은 멀티미디어 카메라 칩 프로젝트를 포함해 팹리스 칩 연구개발과 관리를 오랫동안 해 왔다. 칩 팀이 새 코덱에 대해 던지는 질문은 BD-rate에 관한 것이 아니다. 질문은 이렇다. 픽셀당 몇 줄(J)인가, 픽셀당 DRAM 트래픽은 몇 바이트인가, 게이트는 몇 개인가, 그리고 세상의 모든 디코더가 정확히 같은 비트를 계산할 수 있는가? 네 가지 모두에서 2026년의 신경망 디코더는 첫 칩이 나오기 전의 하드웨어 HEVC와 같은 위치에 있다. 소프트웨어에서는 유망하지만 실리콘에서는 비싸다.

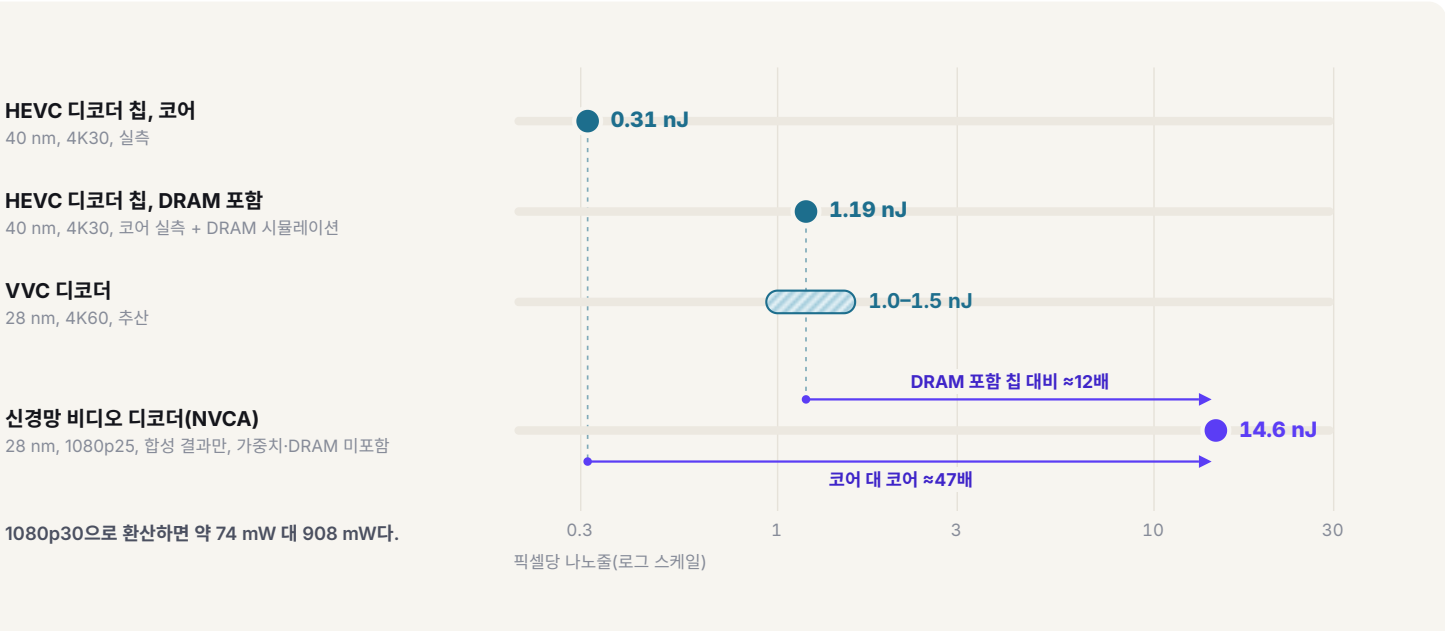
픽셀당 에너지: 10배 수준의 격차

칩으로 설계된 신경망 디코더는 드물어서, 비교는 공개된 설계 몇 건에 기댈 수밖에 없다. 전통 방식 쪽 기준은 MIT의 HEVC 디코더 칩(Tikekar 외, ISSCC 2013 및 IEEE JSSC 2014)이다. 40 nm 공정, 1.77 mm² 면적에 200 MHz로 4K를 초당 30프레임으로 디코딩한다. 코어 소비 전력은 78 mW로 픽셀당 0.31 nJ에 해당하며, 시뮬레이션한 DRAM 전력까지 포함하면 시스템 전체로는 픽셀당 1.19 nJ에 이른다. VVC의 양산 수치는 공개되지 않는다. Electronic Design의 2026년 분석은 MIT 칩을 28 nm로 스케일링하고 VVC의 더 높은 복잡도를 적용해, 약 100만 게이트 규모의 4K60 디코더를 픽셀당 1–1.5 nJ 수준으로 추산한다.

신경망 쪽에서는 Electronic Design의 집계로 유일하게 공개된 엔드투엔드 신경망 비디오 디코더 ASIC 설계가 NVCA(Zhang 외, DATE 2024)다. TSMC 28 nm로 합성되었고 게이트 수는 501만 개이며, 칩 밖에 있는 가중치는 포함하지 않았다. 760 mW로 1080p를 초당 25프레임으로 디코딩하며, 이는 픽셀당 14.6 nJ에 해당한다.

그림 06 **디코딩 픽셀당 에너지: 고정 기능 디코더가 10배 수준 앞선다**

유일하게 공개된 신경망 비디오 디코더 칩 설계는 픽셀당 에너지가 DRAM까지 포함해 실측한 HEVC 디코더 칩의 약 12배, 코어만 놓고 보면 약 47배다. 신경망 설계 자체의 가중치·DRAM 트래픽은 포함하지 않았다.



출처: Tikekar 외, IEEE JSSC 2014(MIT); NVCA, Zhang 외, DATE 2024(arXiv 2312.10716); VVC 범위는 Electronic Design의 추산(2026). HEVC 칩은 더 오래된 공정(40 nm 대 28 nm)이어서 신경망 설계에 유리한 비교다.

이는 DRAM을 포함한 HEVC 칩의 약 12배, 코어만 놓고 보면 약 47배다. 이 비교는 세 가지 면에서 NVCA에 유리하다. HEVC 칩은 더 오래된 공정 노드이고, NVCA 수치에는 가중치·DRAM 트래픽이 빠져 있으며, NVCA는 실측 실리콘이 아니라 합성 결과다. 1080p30에서는 약 74 mW 대 약 900 mW다. 아무도 눈치채지 못하는 디코더와, 휴대폰의 배터리와 발열 예산에 드러나는 디코더의 차이다.

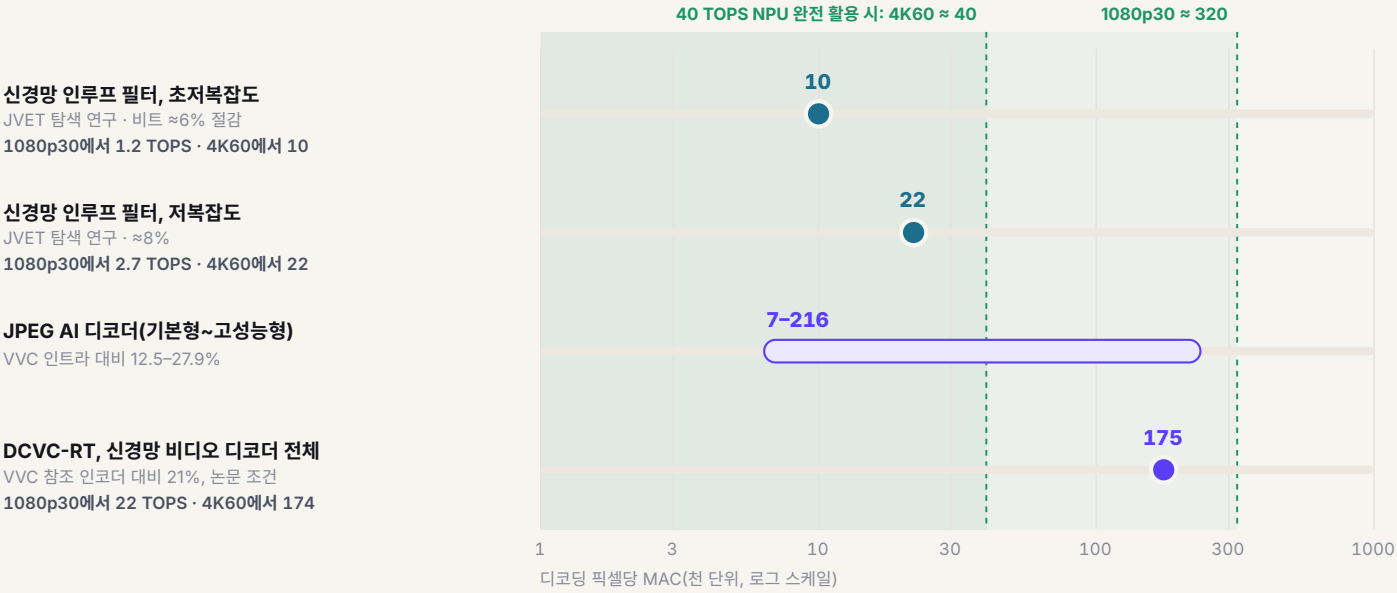
픽셀당 연산량: 신경망 디코더가 NPU에 요구하는 것

필자가 찾은 공개 신경망 비디오 코덱 중 가장 빠른 것은 마이크로소프트 리서치의 DCVC-RT(Jia 외, CVPR 2025)다. 엔비디아 A100 GPU에서 1080p를 초당 125프레임으로 인코딩하고 113프레임으로 디코딩하면서, 논문의 시험 조건에서 VVC 참조 인코더 대비 비트레이트를 21% 절감한다. 또한 서로 다른 기기가 일관되게 디코딩하도록 정수 연산을 쓴다. 보충 자료에는 모듈별 복잡도가 나와 있다. 디코더가 실행해야 하는 네 모듈(디코더, 특징 추출기, 엔트로피 모델, 복원)을 더하면 디코딩 픽셀당 약 17만 5천 회의 곱셈-누산이 나온다 추정.

1080p30에서는 초당 약 11조 회의 곱셈-누산이며, MAC당 2연산으로 치면 활용률 손실을 감안하기 전에 약 22 TOPS다. 4K60에는 그 8배가 필요하다 추정. 비교하자면 마이크로소프트는 AI PC를 최소 40 TOPS의 NPU로 정의하고, 애플은 M4의 Neural Engine을 38 TOPS로 평가한다. 이는 최대치이며, 기기가 하는 다른 모든 일과 나눠 써야 한다.

그림 07 **신경망 도구는 픽셀당 수만 MAC, 신경망 디코더 전체는 수십만 MAC이 든다**

픽셀당 디코더 곱셈-누산(MAC) 수(로그 스케일)를, 40 TOPS NPU를 완전히 활용할 때 4K60과 1080p30에서 감당할 수 있는 수준과 비교했다. 각 이득은 서로 다른 기준 대비이므로 직접 비교할 수 없다.



출처: JVET(SC 29 의장 자료, 2026년 4월); JPEG 위원회(2024); DCVC-RT, CVPR 2025 보충 자료(디코더 모듈을 저자가 합산); 마이크로소프트의 Copilot+ PC 정의. 저자 계산: MAC당 연산 두 번, 완전 활용 가정.

그림 7의 사다리는 이 분야를 읽는 유용한 방법이다. JVET의 인루프 필터(1만 MAC에서 약 6%, 2만 2천 MAC에서 8% 절감)처럼 픽셀당 수만 MAC이 드는 신경망 도구는 미래 표준 디코더 안에 들어갈 만하다. JPEG AI는 이 범위 전체에 걸쳐 있어, 픽셀당 7천에서 21만 6천 MAC으로 VVC 인트라 대비 12.5–27.9%의 이득을 낸다. 픽셀당 20만 MAC 안팎이 드는 신경망 비디오 디코더 전체는 NPU에서, 높지 않은 해상도로, 폐쇄 루프 안에서 돌려야 한다. MLVC가 있는 곳이 정확히 거기다. 540p30에서 NPU를 절반 이하로만 쓴다.

진짜 한계는 MAC이 아니라 메모리 트래픽인 경우가 많다

DCVC-RT 저자들은 속도를 제한한 것이 산술 연산이 아니라 “운용 비용(operational cost)”, 즉 메모리 입출력과 함수 호출 횟수라는 것을 발견했다. 명시적인 모션 모듈과 멀티스케일 처리를 없앤 것도 주로 그 때문이다. NVCA 설계자들은 모션 특징과 잔차 특징에서 생기는 칩 외부 메모리 트래픽을 줄이기 위해 특수한 데이터플로를 만들었다. 칩 설계자라면 이 패턴을 알아볼 것이다. MIT HEVC 칩에서는 DRAM 대역폭을 67% 줄인 캐시를 넣은 뒤에도 DRAM이 시스템 에너지의 약 4분의 3(픽셀당 1.19 nJ 중 0.88 nJ)을 차지한다 분석. 프레임 사이에 큰 다채널 특징 맵을 유지하는 신경망 디코더는 메모리 문제를 개선하기는커녕 악화시킨다.

정확성: 모든 디코더가 같은 비트를 계산해야 한다

엔트로피 코딩은 디코더의 확률 모델이 인코더의 것과 정확히 일치할 때만 작동한다. 전통 방식 코덱에서 그 모델은 표 하나 또는 몇 가지 정수 규칙이다. 신경망 코덱에서는 그것이 신경망이며, 커널이 다르거나 누산 순서가 다르거나 NPU가 달라서 두 기기 사이에 부동소수점 차이가 생기면, 디코더가 엉뚱한 심벌을 읽고 그 지점부터 스트림을 잃을 수 있다.

신경망 코덱을 출시하거나 표준화한 이들은 모두 이 문제를 공학적으로 없애야 했다. DCVC-RT는 기기 간 일관된 코딩을 위해 모델을 정수로 변환한다. MLVC는 플랫폼 간 불일치를 견디기 위해 공유 스케일 하이퍼프라이어 설계를 쓰는데, 이것이 DCVC-RT 대비 압축 효율을 내주는 이유의 일부일 가능성이 크다 분석. JVET의 신경망 탐색 연구는 모든 신경망을 16비트 정수 추론을 지원하는 작은 라이브러리 SADL로 돌린다. JPEG AI는 잠재, 잔차, 픽셀 영역에 적합성 검증 지점을 정의한다. 칩 팀에게 이는 HEVC 정수 변환과 같은 규율이다. 표준이 산술 연산을 정확히 정의해야 하며, 이는 신경망 설계를 제약하고 업체별 지름길을 배제한다.

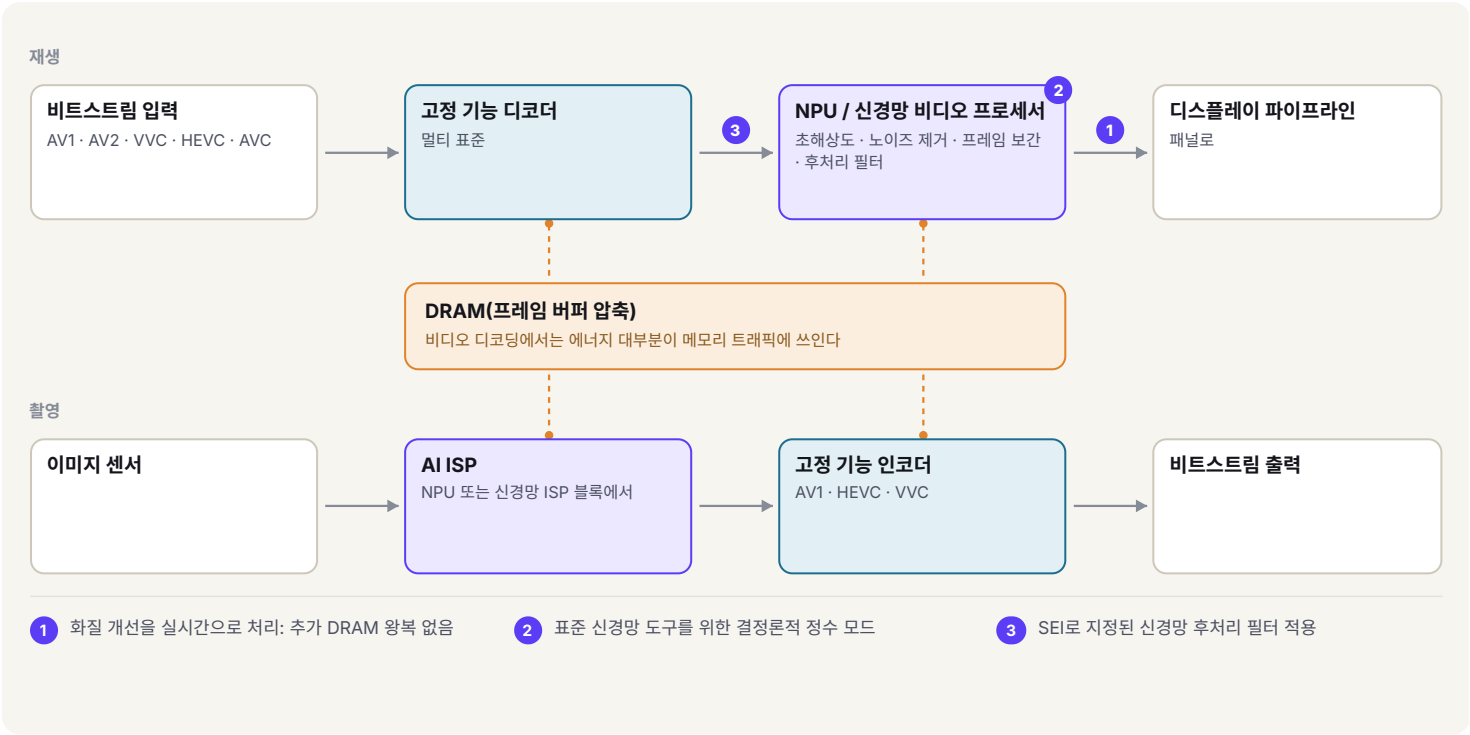
걸림돌은 면적이 아니라 검증과 수명이다

추산된 VVC 디코더의 약 100만 게이트에 비해 NVCA의 500만 게이트는 커 보이지만, 이미 수 TOPS급 NPU를 탑재한 최신 휴대폰이나 TV SoC에서 수백만 게이트는 설계를 좌우하는 요소가 아니다. 실제로 큰돈이 드는 것은 적합성 시험 세트에 대한 검증, 기기 수명 내내 소프트웨어 업데이트를 거치면서도 프로그래머블 디코더가 올바르게 동작하도록 유지하는 일, 그리고 4K60 HDR에서의 전력이다. 이런 비용은 무엇을 만들어야 하는지 표준이 확정한 뒤에야 감당할 의미가 생긴다.

지금 이기는 아키텍처: 고정 기능 코덱 + NPU

그림 08 **상용화된 분업 구조: 비트스트림은 고정 기능 하드웨어, 픽셀은 신경망 엔진**

2026년의 미디어 SoC. 코덱 블록은 모든 표준을 픽셀당 1나노줄 미만으로 디코딩하고, NPU는 신경망이 잘하고 표준이 제약하지 않는 일을 맡는다. 중요한 비용은 DRAM 트래픽이다.



출처: 텐센트(Canghai V2), 칩스앤미디어, Allegro DVT, 삼성전자의 제품 공개 자료(2026)를 바탕으로 한 저자의 아키텍처 관점.

2026년에 팔리는 실리콘은 고정 기능 멀티 표준 코덱과 그 옆의 신경망 엔진을 짝짓는다. 코덱 블록은 AV1, AV2, VVC, HEVC, AVC를 픽셀당 1나노줄 미만으로 처리한다. 신경망 엔진은 신경망이 잘하고 표준이 제약하지 않는 일을 맡는다. 촬영 쪽에서는 노이즈 제거와 ISP, 디스플레이 쪽에서는 초해상도, 프레임 보간, 화질 개선이다. 둘은 프레임 버퍼 압축과 DRAM 스케줄을 공유한다. 중요한 비용이 메모리 트래픽이기 때문이다.

시장이 이 분업을 보여 준다. 텐센트가 2019년부터 자체 개발한 비디오 코덱 ASIC Canghai(沧海)는 1세대 칩 10만 개 이상이 텐센트 클라우드와 자사의 라이브, 주문형, 클라우드 게임 서비스 전반에 배치되어 있다. 2026년 5월 파워온을 마치고 하반기 서비스를 위한 양산에 들어간다고 발표된 2세대는 하나의 아키텍처에서 H.265와 H.266을 지원하며, 이전 세대보다 압축 효율이 각각 10% 이상, 30% 이상 높고, 처리량은 약 2배이며, 초해상도 같은 온칩 화질 개선 기능을 갖췄다. 칩스앤미디어는 코덱 IP와 비디오용 NPU인 WAVE-N을 나란히 판다. Allegro DVT는 AV2 디코더를 팔면서 첫 신경망 비디오 처리 IP를 추가했다. 삼성전자의 TV 프로세서는 표준 디코더와 수백 개의 업스케일링 신경망을 결합한다.

역할별로 필자라면 무엇을 만들까

SoC·TV 칩 팀. 재생해야 하는 모든 표준의 고정 기능 디코더를 유지하라. 지금은 AV2가, 방송이 요구하는 곳에서는 VVC가 여기에 포함된다. NPU는 디스플레이 해상도와 프레임 레이트의 후처리에 맞춰 크기를 정하고, 디코더에서 NPU를 거쳐 디스플레이로 가는 경로를 화질 개선이 DRAM 왕복을 추가하지 않도록 설계하라. 신경망 후처리 필터 SEI를 지원하라. NPU에 결정론적 정수 모드를 제공하라. H.267은 2029년 이후에 설계를 시작할 대상으로 보고, 그 신경망 도구를 위해 어느 정도의 프로그래머빌리티를 남겨 두라.

IP 업체. 지금은 AV2와 VVC 디코더를 팔고, 그 주변에서 신경망 처리에 필요한 부품도 함께 팔아라. 후처리 필터 가속기, 프레임 버퍼 압축, 적합성 도구다.

비디오 플랫폼. 인코더 쪽 신경망 전처리가 가장 먼저 투자를 회수한다. 타이틀당 한 번만 실행되고 기기에 요구하는 것이 없기 때문이다. 기기가 허용하는 곳에서는 클라이언트 측 초해상도를 써라. 엔드투엔드 신경망 코덱은 직접 통제하는 폐쇄 루프에만 써라.

09장

09

표준과 특허

표준과 특허: 신경망 도구가 라이선스 대상이 되는 곳

표준화 파이프라인

표준	기구	신경망 요소	현황(2026년 11월)
H.267(가칭, VVC 이후 세대)	JVET(ITU-T와 ISO/IEC)	미정. 전통 방식 도구(ECM)와 신경망 도구(인루프 필터, 인트라 예측, 초해상도) 모두 후보	2026년 7월 제안 요청, 2027년 1월 응답 평가, 2027년 하반기 첫 테스트 모델, 2029년 말 버전 1
VVC·HEVC(VSEI 포함)	JVET	신경망 후처리 필터를 SEI 메타데이터로 기술할 수 있으며, 적용 여부는 디코더의 선택	시행 중, 2025년 10월 판에서 메시지 확장
AV2	AOMedia	전통 방식 도구, 로열티 무료	2026년 5월 28일 버전 1.0.0, 2026년 6월부터 디코더 IP 라이선스, 유튜브 웹 스트리밍은 2027년 초 계획
JPEG AI	JPEG(ISO/IEC와 ITU-T)	사람과 기계 비전을 위한 엔드투엔드 학습 기반 이미지 코덱	2025년 파트 1 발행, 두 번째 버전 논의 중, 제품의 기본 지원은 확인되지 않음
IEEE 1857.11	IEEE	NPU용으로 설계된 신경망 이미지 코딩	2024년 12월 발행
Video Coding for Machines	MPEG	시청이 아닌 기계 분석을 위한 코딩	2026년 초 국제 표준안, 2026년 말 최종 확정 계획
Opus ML 확장	IETF	선택형 딥 리던던시, 디코더 측 음질 개선, 대역폭 확장	libopus 1.5와 1.6에 포함, 딥 리던던시는 표준화 중

두 가지가 눈에 띈다. 첫째, 신경망 요소는 핵심보다 가장자리에서 먼저 표준에 들어온다. 선택형 후처리 필터, 선택형 화질 개선, 이미지 코딩으로 들어오며, 모두 주 디코딩 경로는 바꾸지 않는다. 둘째, 모든 화면의 디코딩 경로에 신경망 도구를 넣을 수 있는 유일한 표준인 H.267은 첫 버전까지 3년이 남았다.

특허에서 누가 자리를 잡았나

표준 도구는 라이선스 가능한 특허가 되므로, H.267 과정은 특허 경쟁이기도 하며 신경망 도구도 그 일부다. GreyB가 VVC 이후 세대 특허 3,916건을 분석한 2025년 특허 지형 보고서는 바이트댄스를 선두로 꼽는다. 바이트댄스는 특허 패밀리 645건을 보유하고 JVET 기고 251건을 제출했다. GreyB는 샤오미와 시안전자과기대처럼 20건 이상의 패밀리를 보유한 산학 협력 사례도 언급한다. MLex는 2026년 3월 텐센트, 알리바바, 바이트댄스가 비디오 코덱 특허 풀의 적극적인 참여자로 부상하고 있다고 보도했다. 2026년 5월 Sauvegarder Investment Management는 알리바바로부터 AV1, AV2, Video Coding for Machines를 아우르는 비디오 코덱 특허 자산 126건을 매입했다. 2025년 10월 인터디지털은 Deep Render를 인수하면서 Deep Render의 AI 비디오 코딩 특허를 포트폴리오에 더했다.

기기당 금액은 특허 풀 요율표에서 쉽게 확인할 수 있다. Access Advance의 VVC 요율(2026년 7월 1일 기준, 계약을 준수하고 상표를 사용하는 라이선시 대상)은 미국, EU, 일본 같은 Region 1 시장에서 휴대폰, 태블릿, 노트북 대당 0.50달러, Region 2에서는 0.25달러다. 4K TV는 Region 1에서 가격에 따라 1.00달러에서 1.50달러이며, 기업별 연간 상한은 7,500만 달러다. AV1과 AV2는 AOMedia 회원사들이 로열티 없이 제공한다. H.267에 무엇이 들어가든 그 도구들도 같은 방식으로 가격이 매겨질 것이며, 신경망 코딩 특허를 가진 기업들이 지금 JVET에 기고하는 이유가 여기에 있다 분석.

10장

10 누가 돈을 버는가, 기업은 어디서 기회를 찾을 수 있는가

2026년의 돈

그림 09 2026년 신경망 코딩으로 누가 돈을 버는가

돈은 표준과 플랫폼을 따라간다. 엔드투엔드 신경망 비디오 코덱 라이선스로 생계를 꾸리는 독립 기업은 찾지 못했다.

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)

출처: 보고서에 인용한 기업·특허 풀 공개 자료. 2026년 11월 12일 기준. 풀 요율: Access Advance, VVC, Region 1, 계약 준수 및 상표 사용 조건, 2026년 7월 1일 기준.

플랫폼은 절감분을 가져간다. 넷플릭스, 유튜브, 메타, 마이크로소프트, 텐센트는 신경망 처리를 자체 개발하고, 그 이익을 대역폭 절감, 저장 공간 절감, 비트당 화질 향상으로 가져간다. 보급에 도움이 될 때는 코드를 공개한다. 마이크로소프트는 MLVC를 MIT 라이선스로 공개했다.

기기·칩 제조사는 차별화를 판다. 카메라 ISP와 TV 화질 프로세서는 신경망이 모든 사용자에게 닿는 곳이며, 기능으로 판매된다. 퀄컴의 Cognitive ISP, 삼성전자의 업스케일링 프로세서가 그렇다.

IP 업체는 새 표준마다 디코더를, 여기에 더해 NPU를 판다. AV2 규격이 나온 지 3주 안에 칩스앤미디어, Allegro DVT, 베리실리콘이 모두 AV2 디코더 고객을 발표했다. 같은 업체들이 이제 코덱 옆에 신경망 처리 블록도 판다.

GPU 업체는 플랫폼을 만든다. DLSS와 신경망 텍스처 압축은 엔비디아의 텐서 하드웨어를 게임 제작 방식의 일부로 만든다.

특허권자는 기기당 로열티를 걷는다. VVC와 HEVC 특허 풀을 통해 걷고, 2029년 이후에는 H.267에서도 걷게 된다.

신경망 코덱 스타트업은 인수됐다. WaveOne은 2023년 애플에, iSIZE는 2023년 소니 인터랙티브 엔터테인먼트에, Deep Render는 2025년 인터디지털에 인수되었으며, 모두 인수 조건은 공개되지 않았다. 매번 플랫폼이나 라이선서가 팀과 특허를 사들였다. 2026년 기준으로 엔드투엔드 신경망 비디오 코덱 라이선스로 생계를 꾸리는 독립 기업은 찾지 못했다.

기업이 기회를 찾을 수 있는 곳

사업 기회	구매자	지금인 이유	주요 위험
인코더 쪽 신경망 전처리(다운스케일링, 노이즈 제거, 타이틀별 튜닝)	스트리밍·사용자 동영상 플랫폼, 방송사	기기 변경 불필요, 넷플릭스에서 검증	대형 플랫폼의 자체 개발
디스플레이 측 화질 개선 IP와 튜닝(초해상도, 프레임 보간, 화질)	TV·셋톱·휴대폰 칩 제조사	기기에 이미 NPU 탑재, 눈에 보이는 차별화	범용화, 칩 제조사의 자체 개발
폐쇄 루프 신경망 코덱(화상회의, 원격 운용, 전용 링크)	기업 통신·산업용·텔레프레즌스 업체	MLVC가 실현 가능성을 입증했고 코드도 공개	양 끝을 통제해야 함, NPU 파편화
코덱+NPU 실리콘 또는 IP(AV2, VVC, 후처리)	SoC 제조사	AV2 사이클 시작	강력한 기존 강자
평가·적합성·결정론적 추론 도구	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

대부분의 기업에게 2027년에 매출을 낼 수 있는 곳은 처음 두 행이다. 세 번째 행은 한 제품에 집중하는 기업에 가장 흥미로운 기회다. 마지막 행은 인내와 라이선스 전략이 필요하다.

11장

11 주장과 근거
흔히 듣게 될 주장과 근거가 말하는 것

주장	근거가 말하는 것
"신경망 코덱은 이미 VVC를 앞선다."	실험실에서는 그렇다. DCVC-RT는 논문 조건의 1080p에서 VVC 참조 인코더 대비 21%를 절감하고, JPEG AI는 VVC 인트라 대비 12.5–27.9%를 절감한다. 디코더 비용은 픽셀당 약 17만 5천 MAC, 다시 말해 NPU를 돌리는 에너지이며, 고정 기능 하드웨어는 1나노줄에도 못 미친다.
"AI 코덱이 몇 년 안에 H.26x와 AV1을 대체할 것이다."	재생에서는 아니다. 신경망 도구를 실을 수 있는 첫 표준은 2029년 말이 목표이고, 칩은 2031년경, 대중화는 2030년대 중반에 뒤따른다 추정. 지금 도입되는 코덱인 AV2는 전통 방식이다.
"MLVC는 대역폭을 88% 줄인다."	마이크로소프트 자체 테스트셋에서 360p의 H.264 대비 수치다. 하드웨어 HEVC 대비로는 360p에서 75.5%, 540p에서 65.4%이며, 마이크로소프트가 양 끝을 통제하는 통화에서의 결과다.
"신경망 텍스처는 비디오 메모리를 85–96% 줄인다."	시연 장면에서의 이야기다. 출시된 게임은 실효 압축률이 약 30% 높아졌고, 런타임 비용 때문에 일부 에셋에만 신경망 텍스처를 썼다.
"JPEG AI는 ISO 표준이니 카메라와 브라우저가 채택할 것이다."	제품에서 기본 지원은 찾지 못했다. 2026년 웹이 채택한 포맷인 Chrome 155의 JPEG XL은 전통 방식이다.

주장	근거가 말하는 것
“신경망 디코딩은 휴대폰에 너무 비싸다.”	부분적으로 맞다. MLVC는 NPU의 절반 안에서 540p30을 디코딩한다. DCVC-RT급 모델로 4K60을 디코딩하려면 약 175 TOPS의 지속 연산 성능이 필요하다 추정.
“위성·드론 링크가 신경망 비디오의 큰 시장이다.”	기술적으로는 가장 잘 맞지만, 시장이 작고 태평양 양쪽에서 수출 통제가 강화되고 있다.
“대형 플랫폼이 스타트업에서 신경망 코덱을 라이선스할 것이다.”	직접 만들었거나(넷플릭스, 마이크로소프트, 텐센트) 스타트업을 사들였다(애플, 소니, 인터디지털).

간단한 시험 하나로 대부분의 홍보 문구를 걸러 낼 수 있다. 기준, 지표, 해상도, 디코딩 비용을 물어라. 네 가지를 모두 밝히지 못하는 주장은 제품이 아니라 연구 결과다.

12장

12 이정표 2027년부터 2030년까지 주목할 점

시기	이정표	중요한 이유
2027년 1월	JVET, 브리즈번에서 H.267 응답 평가	상위 응답 중 몇 건이 신경망 도구에 의존하는지, 픽셀당 비용은 얼마인지
2027년 초	유튜브의 첫 AV2 웹 스트리밍, 발표된 SoC의 AV2 디코더	다음 전통 방식 사이클이 2030년 무렵까지 기기 예산을 정한다
2027년 하반기	H.267의 첫 테스트 모델	신경망 도구가 핵심 설계에 들어가는지, 탐색 연구에만 머무는지
2027년	Teams P2P 통화 너머로의 MLVC 확산, 공개 코드를 채택하는 다른 업체	공개 신경망 코덱이 소유자 밖에서 치르는 첫 시험
2027–2028년	JPEG AI를 기본으로 쓰거나 읽는 첫 제품	기업 경계를 넘어 쓰이는 첫 신경망 포맷
2027–2028년	정식 출시 드라이버의 DirectX Linear Algebra, 더 많은 게임과 모바일의 신경망 텍스처	그래픽 안에서의 업체 간 호환 신경망 디코딩
상시	휴대폰과 PC의 NPU 예산	지속 성능 약 100 TOPS라면 DCVC-RT급 1080p30 디코더가 NPU의 절반이 아니라 약 5분의 1을 차지한다 분석
2029년 말	H.267 버전 1	대중적 재생에 쓰일 신경망 도구의 실리콘 사이클 시작

13장

13 열린 질문

열린 질문

이 보고서의 몇몇 질문에는 아직 공개된 답이 없으며, 필자가 기술적으로 가장 흥미롭게 여기는 것도 바로 이 질문들이다.

> **대규모 정확성.** 신경망 디코더는 압축 우위의 대부분을 잃기 전까지, 서로 다른 업체의 NPU에서 비트 단위로 일치하는 정수 전용 연산에 얼마나 가까이 갈 수 있는가? MLVC가 내준 6%포인트는 하나의 데이터 포인트일 뿐이며, 표준에는 일반적인 답이 필요하다.

> **MAC보다 메모리.** 하드웨어 디코더와 디스플레이 사이에서 실시간으로 실행되는 신경망 후처리는 픽셀당 DRAM 트래픽이 얼마나 들며, 작업 세트를 얼마나 작게 만들 수 있는가?

> **H.267에서 신경망 도구의 값.** 4K60 디코더에서 신경망 인루프 필터는 픽셀당 몇 MAC, 몇 나노줄부터 절감하는 비트만큼의 값을 하지 못하게 되는가?

> **통화 너머의 폐쇄 루프.** 클라우드 사진 저장과 게임 스트리밍부터 로봇과 전용 중계 링크까지, 다른 어떤 폐쇄형 시스템이 H.267이 나오기 전에 엔드투엔드 신경망 코덱을 채택할 수 있는가?

> **생성하는 디코더.** 비디오 VAE가 이미 그렇듯 디코더가 디테일을 복원하는 대신 생성할 때, 충실도는 어떻게 측정하고 규정해야 하는가?

앞으로 몇 년간 기술 교류와 기술 협력을 통해 함께 탐구할 수 있는 분야들이다. 연락처: i@tanshuai.com

이 보고서의 한계

이 보고서는 2026년 11월 12일까지의 공개 문서에 기반한 문헌 조사다. 이번 판을 위해 직접 인터뷰나 독립 측정은 하지 않았다. MLVC의 비트레이트 절감률이나 신경망 텍스처 압축률처럼 몇몇 핵심 수치는 업체나 저자가 자체 테스트셋에서 얻은 주장이며, 그렇게 표시했다. DCVC-RT에 대한 필자의 연산량 추정은 저자들이 공개한 모듈별 수치를 더한 것으로 활용률 손실을 무시했으므로, 실제 NPU 부하는 더 높을 것이다. 에너지 비교는 서로 다른 공정 노드의 공개 설계 몇 건에 기댄다. 양산 VVC·AV2 디코더 수치는 공개되지 않는다. 표준화 일정은 위원회 자체의 목표다. 2029년 이후의 H.267 일정은 AV1과 VVC의 이력을 바탕으로 한 필자의 추정이다.

저자 소개

탄 슈아이는 칩, 하드웨어, 시스템 소프트웨어 분야에서 20년 넘게 일해 온 기술 전문가이자 창업가이며, 등록 특허 20건을 보유하고 있다. 풀 커스텀 ASIC 데이터패스 설계, 비디오 코덱 가속기 같은 멀티미디어 카메라 칩 프로젝트를 포함해 팹리스 칩 연구개발과 관리 경험을 오랫동안 쌓았으며, 이 보고서도 바로 그 관점에서 썼다. 픽셀당 에너지, 메모리 트래픽, 그리고 표준이 디코더의 어떤 부분을 실리콘에 고정할 수 있게 해 주는가 하는 관점이다. 실리콘밸리, 뉴욕, 시카고, 대만, 그레이터 베이 에어리어(선전과 홍콩), 베트남, 말레이시아를 거치며 일했다. 현재는 AI 연구자이자 개발자로 일하고 있다.

기술 논의: i@tanshuai.com · tanshuai.com

일러두기 및 면책 고지

이 보고서는 2026년 11월 12일까지 공개된 문서에 기반한 독립적인 문헌 조사다. 정보 제공만을 목적으로 하며 투자, 법률, 엔지니어링 자문이 아니다. 제품, 표준화 일정, 특허 풀 요율, 수출 규정은 바뀐다. 근거로 삼는 내용은 반드시 1차 출처와 대조해 확인하라. 업체 주장, 추정, 분석으로 표시한 수치는 그 성격에 맞게 읽어야 한다. 기업명과 제품명은 각 소유자의 것이며 식별 목적으로만 사용했다. 이 보고서를 검토하거나 후원한 기업은 없다.

© 2026 탄 슈아이. 출처를 밝히고 내용을 바꾸지 않는 조건으로 이 보고서를 공유할 수 있다.

출처

모두 2026년 11월 3일부터 11월 12일 사이에 접속했다. 장별로 분류했다.

트래픽, 기기, 범위

- Ericsson, *Mobility Report*, 2026년 6월(2025년 말 모바일 데이터 트래픽 중 비디오 약 75%): https://www.ericsson.com/en/reports-and-papers/mobility-report/reports/june-2026 ; 요약: 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, 전 세계 스마트폰 출하량, 2025년 4분기 및 연간(12억 6천만 대), 2026년 1월 13일: https://www.idc.com/resource-center/press-releases/q425mobilephonetop5/
- TrendForce, 2025년 TV 출하량 전망(1억 9,571만 대): 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, 2025년 TV 전망(2억 870만 대): https://iconnect007.com/article/145954/omdia-forecasts-flat-tv-shipments-in-2025-but-mini-led-poised-for-explosive-growth-at-829/145951/ein

재생과 스트리밍

- Netflix Technology Blog, “AV1 — Now Powering 30% of Netflix Streaming”, 2025년 12월: 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”, 2022년 11월: 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, 2026년 5월 28일: 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”(유튜브 카탈로그 중 AV1 비중): 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”(DTV+ 상업 운영, 2026년 6월): https://mpegh.com/brazil/ ; TV 내장 수신기에 관한 CartaCapital 기사: https://www.cartacapital.com.br/?p=628337 ; TV 3.0의 VVC에 관한 Olhar Digital 기사: 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”, 2025년 10월 29일: 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, 2025년 10월 16일: https://atscaleconference.com/how-meta-deployed-super-resolution-at-scale-to-transform-video-quality/
- Samsung, 8K TV 주요 특징(NQ8 AI Gen3, 신경망 768개): https://www.samsung.com/us/tvs/8k-tv/highlights/

사진과 이미지

- 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 및 Visionary.ai, 완전 AI 기반 ISP, 2026년 1월 5일: 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/ ; PetaPixel 인터뷰: https://petapixel.com/2026/01/18/a-look-at-the-worlds-first-full-ai-based-image-signal-processor
- JPEG committee, 제103차 회의 보도자료(JPEG AI 이득, 복잡도, 스마트폰 구현): https://jpeg.org/items/20240507_press.html
- G. J. Sullivan, ISO/IEC JTC 1/SC 29 소개, 2026년 4월(JPEG AI 현황, ECM·NNVC 수치): 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/
- IEEE 1857.11 작업반(신경망 기반 이미지 코딩): https://sagroups.ieee.org/fvc
- Chrome 155와 JPEG XL, 2026년 10월 6일: https://xenospectrum.com/en/chrome-155-jpeg-xl-rust/

통화와 회의

- Microsoft, “Satin: Microsoft's latest AI-powered audio codec for real-time communications”, 2021년 2월 17일: 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)”, 2026년 7월 24일: https://techcommunity.microsoft.com/blog/linuxandopensourceblog/announcing-the-open-source-release-of-ml-video-codec-mlvc/4539875 ; 코드: 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 ; iReplay 분석: https://ireplay.tv/blog/mlvc-microsoft-neural-video-codec-open-source-streaming/
- Opus 1.5 릴리스 노트: 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”, 2024년 6월 13일: https://engineering.fb.com/2024/06/13/web/mlow-metas-low-bitrate-audio-codec/
- Microsoft, Copilot+ PC(40+ TOPS NPU): https://www.microsoft.com/en-us/windows/copilot-plus-pcs ; Apple, M4 발표(38 TOPS Neural Engine): https://www.apple.com/newsroom/2024/05/apple-introduces-m4-chip/

생성형 AI

- Kyutai, Mimi 모델 카드(12.5 Hz, 1.1 kbps): https://huggingface.co/kyutai/mimi
- Lightricks, “LTX-Video: Realtime Video Latent Diffusion”(1:192 비디오 VAE): https://arxiv.org/abs/2501.00103
- MPEG 제153차 회의 보도자료(Video Coding for Machines, DIS 단계): https://www.mpeg.org/wp-content/uploads/2026/03/press_release_153.pdf

게임과 그래픽

- NVIDIA, DLSS 4 채택 현황: 출시 및 Computex 2025(75개, 125개 이상): https://www.nvidia.com/en-us/geforce/news/125-dlss-4-multi-frame-gen-games-more-announced-computex-2025/ ; Gamescom 2025(175개 이상): https://www.nvidia.com/en-us/geforce/news/gamescom-2025-dlss-4-full-ray-tracing-game-announcements/ ; CES 2026(250개 이상, 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”, 2026년 2월: https://www.ubisoft.com/en-us/studio/laforge/news/415slmB3ZGzv8d2LhO9QT/shipping-neural-texture-compression-in-assassins-creed-mirage
- TweakTown, 엔비디아의 GTC 2026 신경망 텍스처 시연 기사(6.5 GB에서 970 MB로): https://www.tweaktown.com/news/110886/nvidias-neural-texture-compression-cuts-vram-usage-from-6-5gb-down-to-970mb/index.html
- Microsoft DirectX blog, Shader Model 6.9 정식 출시, 2026년 2월 26일: https://devblogs.microsoft.com/directx/shader-model-6-9-retail-and-more/ ; cooperative vector 관련: https://devblogs.microsoft.com/directx/shader-model-6-9-and-the-future-of-cooperative-vector/ ; GDC 2026 계획: https://windowsreport.com/microsoft-showcases-neural-rendering-and-dxr-2-0-plans-for-directx-at-gdc-2026/

특수 링크

- Global Trade Alert, 중국 2026년 제49호 공고: 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(무인 지상 플랫폼 기준치): https://www.armyrecognition.com/news/army-news/2026/china-tightens-drone-export-controls-raising-supply-chain-risks-for-the-us-europe-and-ukraine
- DroneLife, FCC의 외국산 드론·부품 Covered List 등재, 2025년 12월 22일: https://dronelife.com/2025/12/22/fcc-adds-foreign-made-drones-and-components-to-covered-list-citing-national-security-risks/

실리콘

- 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 ; MIT 칩 페이지(4K30에서 78 mW): 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 ; 보충 자료(모듈별 복잡도): https://openaccess.thecvf.com/content/CVPR2025/supplemental/Jia_Towards_Practical_Real-Time_CVPR_2025_supplemental.pdf
- NNVC와 SADL 정수 추론 라이브러리: https://arxiv.org/abs/2309.05846
- Tencent Canghai V2, 2026년 5월 27일(Shanghai Securities News, 10jqka 경유): https://stock.10jqka.com.cn/20260527/c677016020.shtml ; 기능 요약: https://www.besthub.dev/articles/tencent-s-canghai-v2-chip-enters-mass-production-what-the-new-capabilities-mean-for-video-encoding-bcfc18dbfa29
- Chips&Media AV2 디코더 IP, 2026년 5월 12일: https://www.design-reuse.com/news/202530501-chips-media-completes-development-of-next-gen-av2-hw-decoder-ip/ ; AV2 라이선스, 2026년 6월 17일: https://www.edge-ai-vision.com/?p=58270
- Allegro DVT, AV2 디코더 IP, 2026년 6월 16일: https://www.edge-ai-vision.com/2026/06/allegro-dvts-pulsar-decoder-ip-adds-support-for-av2-video-codec/
- VeriSilicon, AV2 지원 VC9800D, 2026년 6월 10일: https://www.businesswirechina.com/en/news/63176.html

표준과 특허

- MPEG 제155차 회의 보도자료(VVC 이후 세대 공동 제안 요청), 2026년 8월 5일: https://www.mpeg.org/wp-content/uploads/2026/08/ag3w254SN26474Press-Release-MPEG155.pdf
- MPEG 제152차 회의 보도자료(증거 제출 요청 결과, VVC·VSEI 새 판), 2025년 10월 30일: https://www.mpeg.org/wp-content/uploads/2025/11/Press-Release-MPEG152.pdf ; ACM SIGMM MPEG 칼럼: 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, VVC 이후 세대(H.267) 특허 지형: https://greyb.com/blog/exclusive-video-codec-landscape/ ; AV2 라이선스와 알리바바 포트폴리오 매각 관련: https://greyb.com/blog/av2-licensing/

- MLex, “Fragmented codec licensing exposes need for clearer rules for Chinese tech majors”, 2026년 3월 27일: https://www.mlex.com/mlex/articles/2458402/fragmented-codec-licensing-exposes-need-for-clearer-rules-for-chinese-tech-majors
- Access Advance, MCBA 및 VVC Advance 로열티 요율(2026년 7월 1일 기준): https://accessadvance.com/wp-content/uploads/2025/07/MCBA-Program-Overview-September-2026.pdf

돈과 기업

- 인터디지털의 Deep Render 인수, 2025년 10월 30일: https://www.advanced-television.com/2025/10/30/interdigital-acquires-deep-render
- TechCrunch, “Apple acquired a startup using AI to compress videos”, 2023년 3월 27일: https://techcrunch.com/2023/03/27/apple-acquired-a-startup-using-ai-to-compress-videos
- Neowin, “Sony acquires AI video startup iSIZE”, 2023년 11월: https://www.neowin.net/news/sony-acquires-ai-video-startup-isize-to-enhance-playstation-streaming/

학습 기반 압축 · 재생 × 실리콘

2026년 신경망 코딩

코덱을 대체하지 않고 코덱 주변에서 상용화된다

이 보고서의 핵심 수치에는 모두 출처와 날짜를 달았다. 저자가 직접 도출한 수치는 '추정'으로, 판매 회사만 발표한 수치는 '업체 주장'으로 표시했다.

탄 슈아이

기술 전문가이자 창업가 · AI 연구자이자 개발자

tanshuai.com · i@tanshuai.com

공개판 · 2026년 11월 12일 기준 데이터