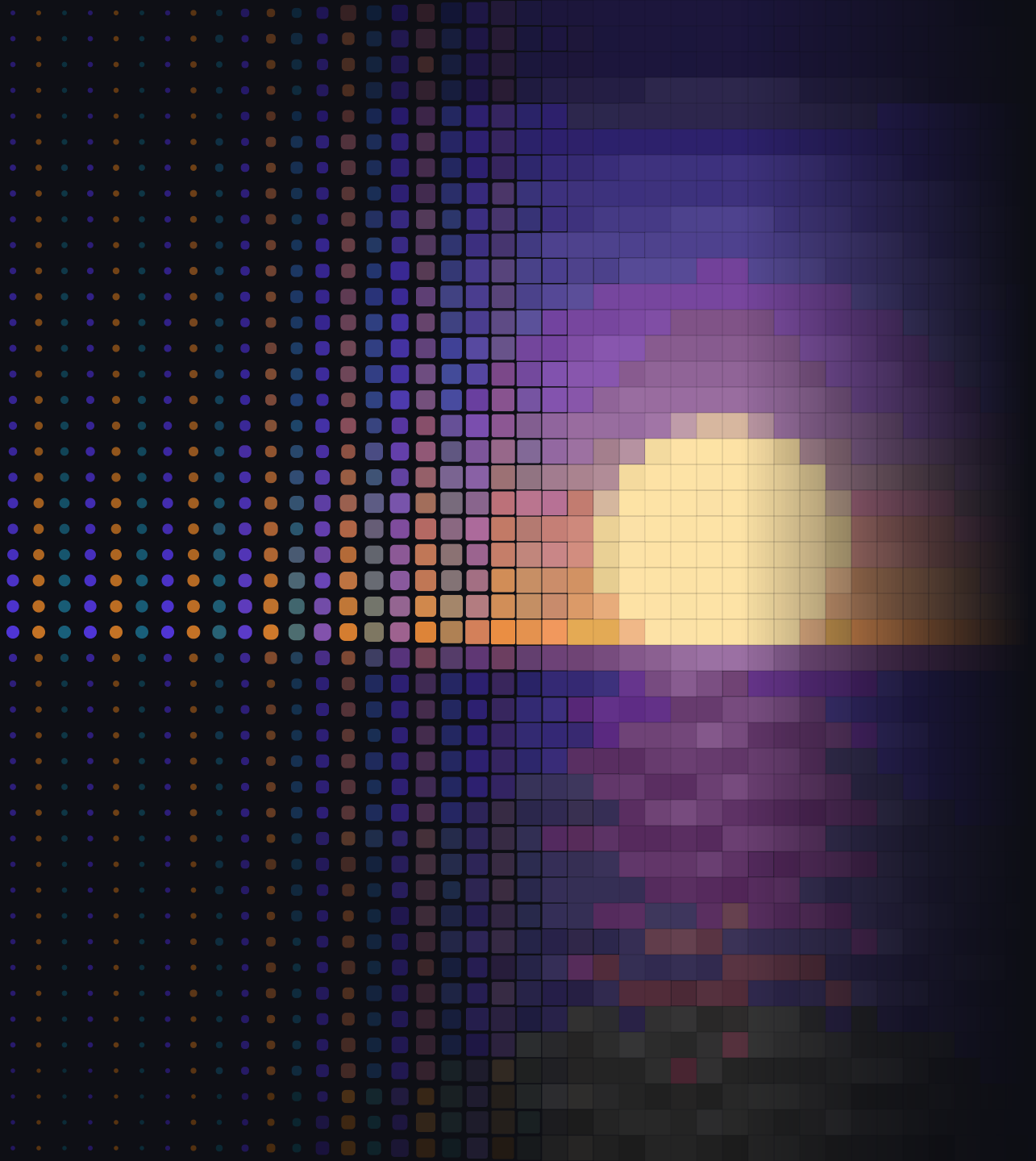


מסמך מחקר · נובמבר 2026

קידוד נוירוני ב־2026

סביב הקודק, לא במקומו

ניגון · צילום · שיחות · בינה מלאכותית יוצרת · משחקים · סיליקון

תאן שואי

tanshuai.com · i@tanshuai.com

כבר בשימוש רחב, אבל סביב הקודק

01 רשתות נוירונים כבר נפוצות בשני הקצוות של הקודקים התקניים

ISP במצלמות, הקטנת רזולוציה בצד המקודד ורזולוציית־על בניגון. זרם הסיביות שביניהם הוא עדיין AV1, HEVC, AVC או VVC, ומפוענח בחומרה ייעודית.

02 קודקים נוירוניים מקצה לקצה נמצאים בשימוש רק במעגלים סגורים

שיחות, טוקנייזרים של בינה מלאכותית יוצרת ומשחק AAA אחד: חברה אחת שולטת בשני הקצוות ויכולה לחזור לקודק תקני.

03 זרמי סיביות נוירוניים במסכים הנפוצים ממתינים ל־H.267

גרסה 1 צפויה בסוף 2029; שבבים ראשונים בסביבות 2031 ושימוש רחב באמצע שנות ה־2030 . אומדן

כ־75%

מתעבורת הנתונים הסלולרית היא וידאו (סוף 2025)

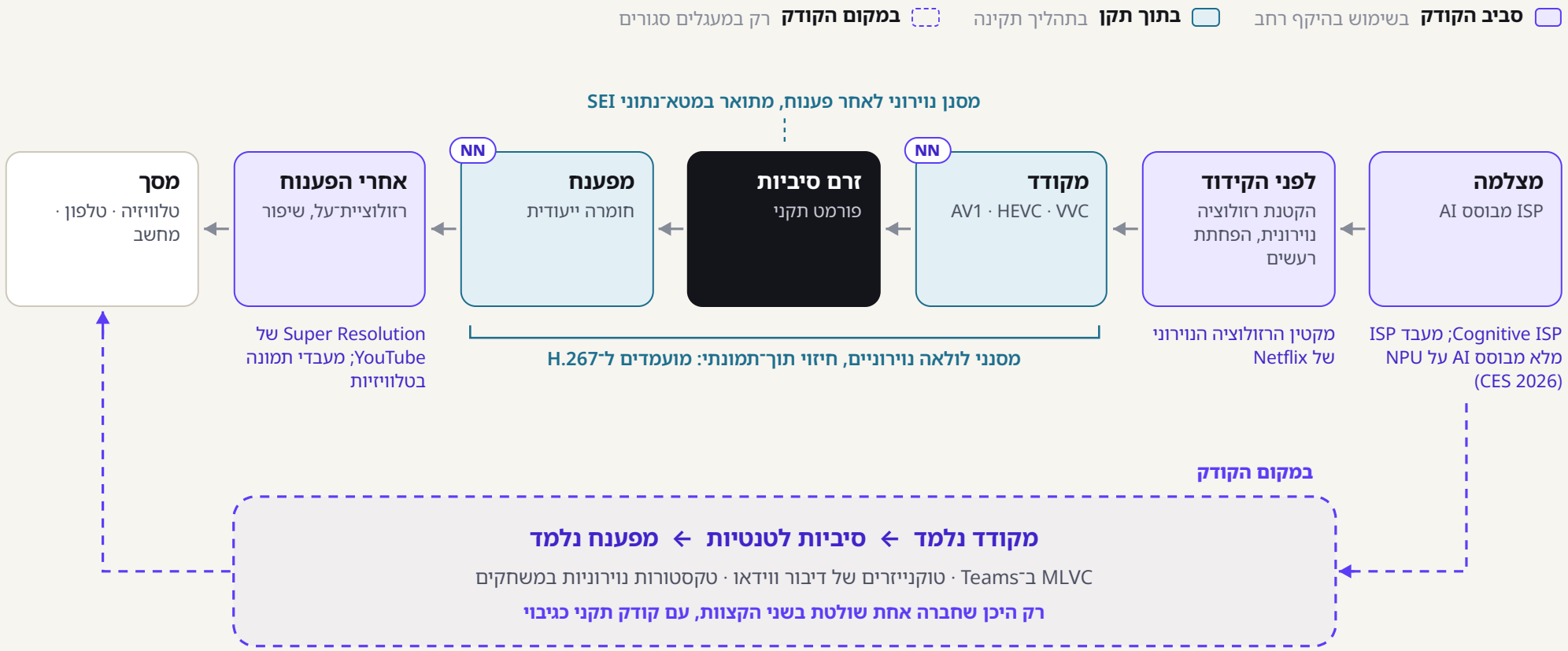
1.26 מיליארד

סמארטפונים שווקו ב־2025, וכולם מפענחים וידאו בחומרה ייעודית

פי 12 בערך

מהאנרגיה לפיקסל של מפענח HEVC בחומרה, בתכנון השבב היחיד שפורסם של מפענח נוירוני

שלוש דרכים שבהן רשת נוירונים נכנסת לקודק



סביב הקודק: לא צריך רשות מאיש. בתוך תקן: צריך תקן. במקום הקודק: צריך שליטה בשני הקצוות.

היכן קידוד נוירוני נמצא בשימוש ב־2026

	סביב קודק תקני עיבוד מקדים, שיפור, הגדלת רזולוציה	קודק נוירוני מקצה לקצה מקודד ומפענח נלמדים	בתוך תקן כלים נוירוניים בזרם סיביות תקני או ב־API
01 ניגון וסטרימינג	**בהיקף רחב** הקטנת רזולוציה ב־Netflix; רזולוציית־על ב־YouTube וב־Meta; הגדלת רזולוציה מבוססת AI בטלוויזיות	**לא בשימוש** אף שירות סטרימינג אינו שולח פורמט וידאו נוירוני; טלוויזיות וטלפונים מפענחים קודקים תקניים בחומרה	**בתהליך תקינה** גרסה 1 של H.267 צפויה בסוף 2029; כלים נוירוניים נמצאים בין המועמדים
02 צילום ודימות	**בהיקף רחב** מעבדי ISP מבוססי AI בטלפוני דגל; ISP נוירוני מלא הוצג כ־IP ב־CES 2026	**הדגמות ופיילוטים** JPEG AI הוא תקן ISO: הדגמות בטלפונים, אין מוצר	**לא בשימוש** הפורמט שצובר תאוצה הוא קלאסי: JPEG XL מופעל כברירת מחדל ב־Chrome 155
03 שיחות ופגישות	**בהיקף רחב** רזולוציית־על בצד המקבל ודיכוי רעשים נוירוני	**בהיקף רחב** שמע נוירוני בשיחות בין שני משתתפים ב־Teams מאז 2021; וידאו MLVC בפריסה ב־Teams	**בשימוש מוגבל** Opus 1.5 מוסיף כלי ML אופציונליים ותואמים לאחור
04 בינה מלאכותית יוצרת	—	**בהיקף רחב** כל מודל וידאו וקול מריץ טוקנייזר נלמד, למשל VAE לווידאו ביחס 1:192	—
05 משחקים וגרפיקה	**בהיקף רחב** הגדלת רזולוציה ויצירת פריימים נוירוניות: DLSS 4 לבדו ביותר מ־250 משחקים ואפליקציות	**בשימוש מוגבל** טקסטורות נוירוניות יצאו לשוק במשחק AAA אחד; ה־SDK של NVIDIA בבטא	**בתהליך תקינה** DirectX מוסיף פעולות מטריצה לשיידרים (תצוגה מקדימה)
06 ערוצים ייעודיים: לוויין, רובוטים	**הדגמות ופיילוטים** שחזור בצד השרת של שידורים בקצב סיביות נמוך	**הדגמות ופיילוטים** קודקים נוירוניים עמידים לאובדן נתונים, במחקר ובפיילוטים	—

השורות מסודרות לפי מספר האנשים שכל תחום מגיע אליהם.

נוירוני בשני הקצוות, תקני באמצע

במקודד

מקטין הרזולוציה הנוירוני של Netflix פועל מאז 2022 בלי שום שינוי במכשירים; כ־77% מהצופים העדיפו אותו בבדיקות של Netflix. טענת החברה

במסך

Super Resolution של YouTube בטלוויזיות מאז אוקטובר 2025; Meta מריצה רזולוציית־על בשרתים שלה ובטלפונים.

בטלוויזיה

Samsung מפרסמת שבמעבד NQ8 AI Gen3 שלה יש 768 רשתות נוירונים להגדלת רזולוציה.

זרם הסיביות נשאר תקני

30%

מהצפייה ב־Netflix היא ב־AV1 (דצמבר 2025)

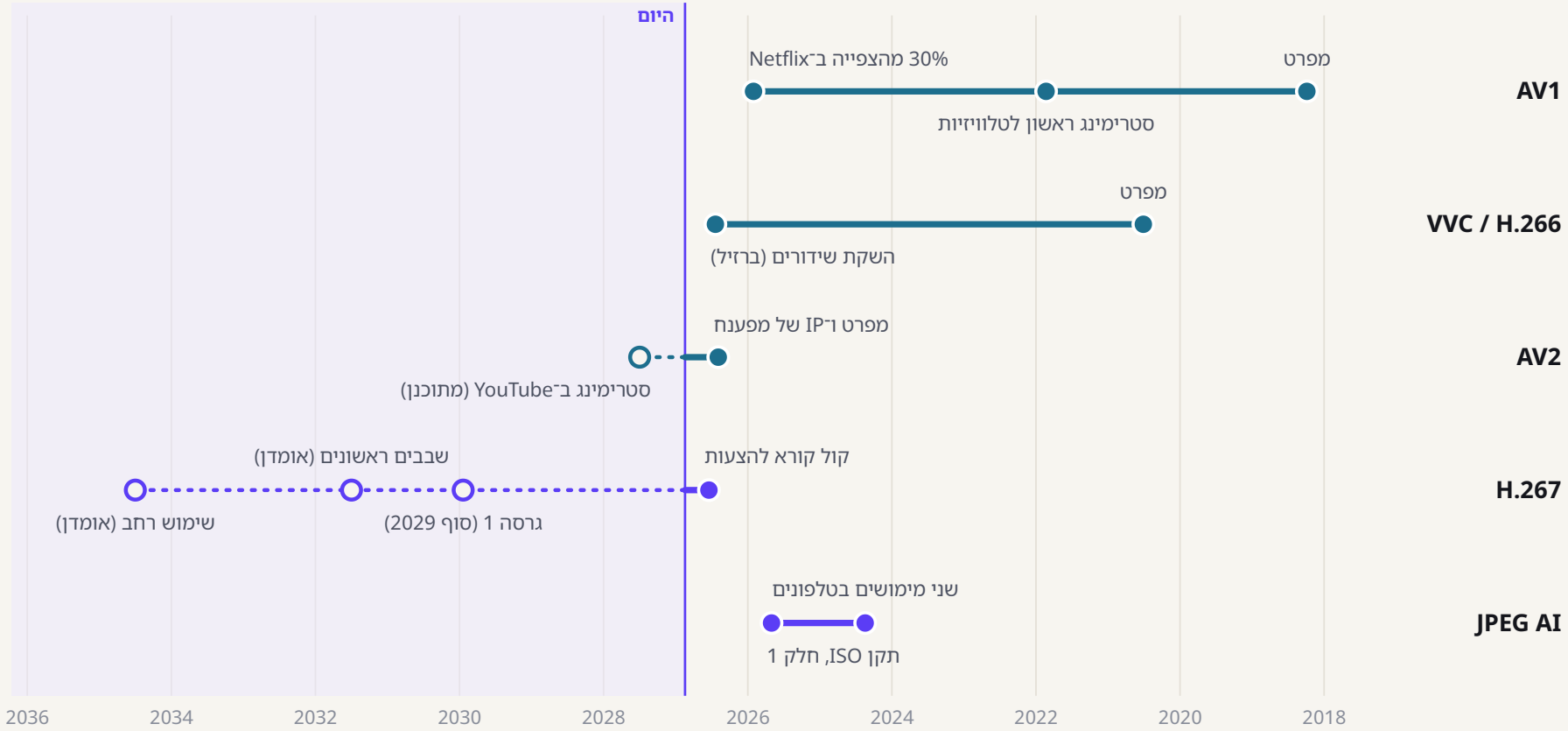
מעל 90%

מהקטלוג של YouTube, משוקלל לפי זמן צפייה, זמין ב־AV1 (מרץ 2026)

3 שבועות

מגרסה 1.0.0 של AV2 (28 במאי 2026) ועד לקוחות ל־IP של מפענח אצל שלושה ספקים

הדרך מתקן גמור למסכים הנפוצים אורכת רוב עשור



ל־AV1 נדרשו כמעט שמונה שנים מהמפרט (2018) ועד 30% מהצפייה ב־Netflix (דצמבר 2025). סימנים חלולים הם יעדים או אומדנים שלי.

H.267 יכריע אם כלים נוירוניים יגיעו לכל מסך

אוקטובר 2025
קול קורא לראיות: חמישה מענים מ־16 חברות ומעבדות; היתרון הנוירוני ברור ביותר ב־8K HDR

יולי 2026
הקול הקורא להצעות פורסם בז׳נבה

ינואר 2027
הערכת המענים בבריסביין

המחצית השנייה של 2027
מודל הבדיקה הראשון

סוף 2029
גרסה 1 של התקן

כ־28%
חיסכון לעומת VVC במסלול הקלאסי ECM

6–14%
חיסכון של מסנני לולאה נוירוניים, בעלות של 10 עד 22 אלף MAC לפיקסל ומעלה

SEI
זרמי VVC כבר יכולים לסמן מסנן נוירוני לאחר פענוח שהמפענחים יפעילו

הקריאה שלי: ליבה קלאסית עם כמה כלים נוירוניים, היכן שהסיביות שנחסכות מצדיקות את עלות המפענח ניתוח.

נוירוני בתוך המצלמה, קלאסי באינטרנט

מצלמה

ה־Cognitive ISP של Qualcomm מריץ סגמנטציה בזמן אמת בטלפונים מאז 2023; ב־CES 2026 הוצג מעבד ISP מלא מבוסס AI שרץ על NPU.

JPEG AI

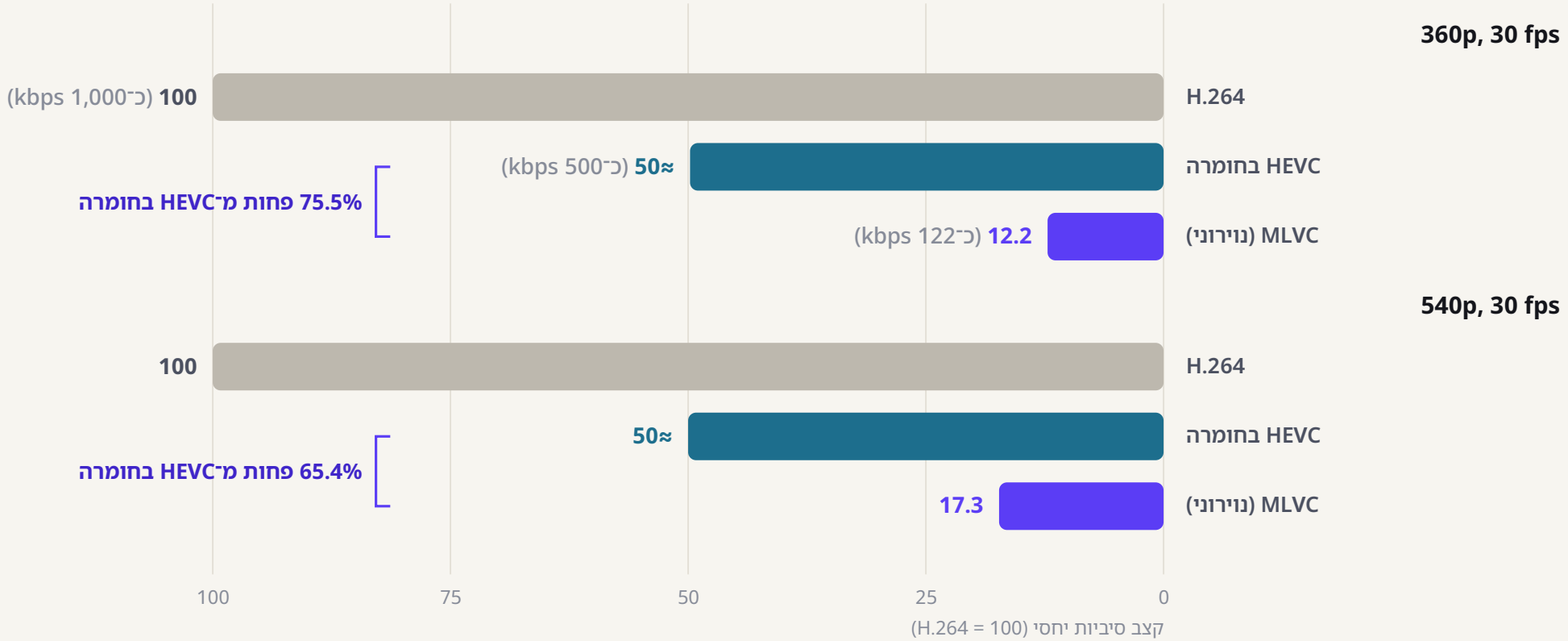
פורסם ב־2025: 12.5–27.9% לעומת VVC תוך־תמונתי, בעלות של 7 עד 216 אלף MAC לפיקסל. לא מצאתי מוצר שמשתמש בו כברירת מחדל.

אינטרנט

Chrome 155 הפעיל את JPEG XL כברירת מחדל ב־6 באוקטובר 2026, עם מפענח בטוח־זיכרון שנכתב ב־Rust.

פורמטים שחייבים להיפתח בכל מקום נשארים קלאסיים; קידוד נוירוני מנצח בתוך צינור העיבוד של חברה אחת.

בשיחות, קודקים נוירוניים מקצה לקצה כבר בשימוש



Satin פועל בכל השיחות בין שני משתתפים ב־Teams וב־Skype מאז 2021. MLVC, בקוד פתוח ברישיון MIT מאז 24 ביולי 2026, נמצא בפריסה בשיחות עמית־לעמית ב־Teams, עד 540p30, ומנצל לכל היותר מחצית מה־NPU. **טענת החברה**

לא כל פלטפורמה עברה לנוירוני, והסיבות מלמדות

בתוך תקן, אופציונלי

Opus 1.5 ו־1.6

שיפור נוירוני, הסתרת אובדן מנות, יתירות עמוקה והרחבת רוחב פס, כולם כבויים כברירת מחדל ותואמים במלואם לתקן מ־2012.

קלאסי, מבחירה

MLow של Meta (2024)

יותר מ־20% מהשיחות של Meta מתבצעות בטלפוני ARMv7, ועשרות מיליוני שיחות WhatsApp ביום בטלפונים בני יותר מעשר שנים. MLow מספק בערך פי שניים מהאיכות של Opus ב־6 kbps, עם 10% פחות חישוב.

תקציב הסיבוכיות של בסיס המכשירים המותקן הוא שמכריע. קודקים נוירוניים נכנסים לשימוש קודם היכן שהמכשירים חדשים וספק אחד שולט בשני הקצוות.

הפריסה הגדולה ביותר, שאף אחד לא קורא לה קודק

- **קודקים נלמדים בכל מקום**

  כל מודל קול ווידאו קורא וכותב את הטוקנים או הייצוגים הלטנטיים של קודק נלמד.

- **לא פורמטים להפצה**

  לכל מודל מרחב לטנטי משלו: אין בקרת קצב, אין בדיקת התאמה, אין מפענח שני.

- **שורת עלות**

  למי שמספק שירותי יצירת וידאו, ייצוג לטנטי דחוס יותר מוריד את עלות היצירה יותר מאשר ביחס ישר ניתוח .

שתי דוגמאות

12.5 Hz · 1.1 kbps

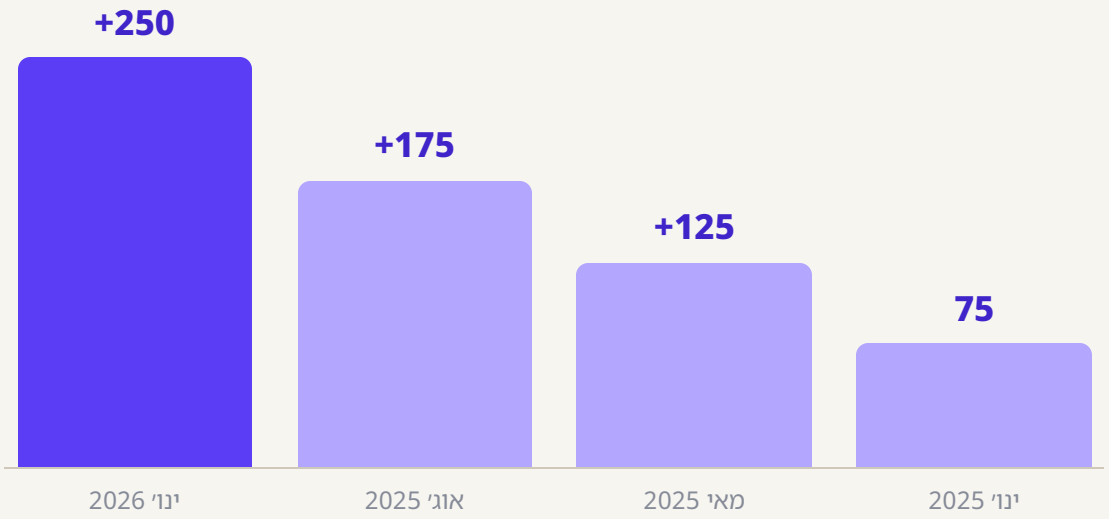
טוקני הדיבור של Mimi מבית Kyutai, מתחת לכל קודק טלפוניה

1:192

ה־VAE לווידאו של LTX-Video: טוקן לטנטי אחד לכל 32×32 פיקסלים לאורך 8 פריימים

שחזור נוירוני כבר נפוץ; נכסים נוירוניים רק מתחילים

DLSS 4 Multi Frame Generation: משחקים ואפליקציות



דחיסת טקסטורות נוירונית

סצנת הדגמה של NVIDIA, GTC 2026

זיכרון טקסטורות: מ־6.5 GB ל־0.97 GB

85% פחות זיכרון

Assassin's Creed Mirage, בגרסה המסחרית

נכסים נבחרים בלבד; לעומת פורמטי בלוקים

דחיסה טובה בכ־30%

מדדים שונים: לא ניתנים להשוואה ישירה

DirectX Linear Algebra נכנס לתצוגה מקדימה ציבורית באפריל 2026: נתיב קוד אחד לפענוח נוירוני בתוך שיידרים, על GPU של כל יצרן.

לווייניים, רחפנים ורובוטים: ההתאמה הטובה ביותר, שוק קטן

למה זה מתאים

בעלים אחד שולט בשני הקצוות, רוחב הפס מועט, והתוכן צר מספיק כדי לאמן עליו.

סין

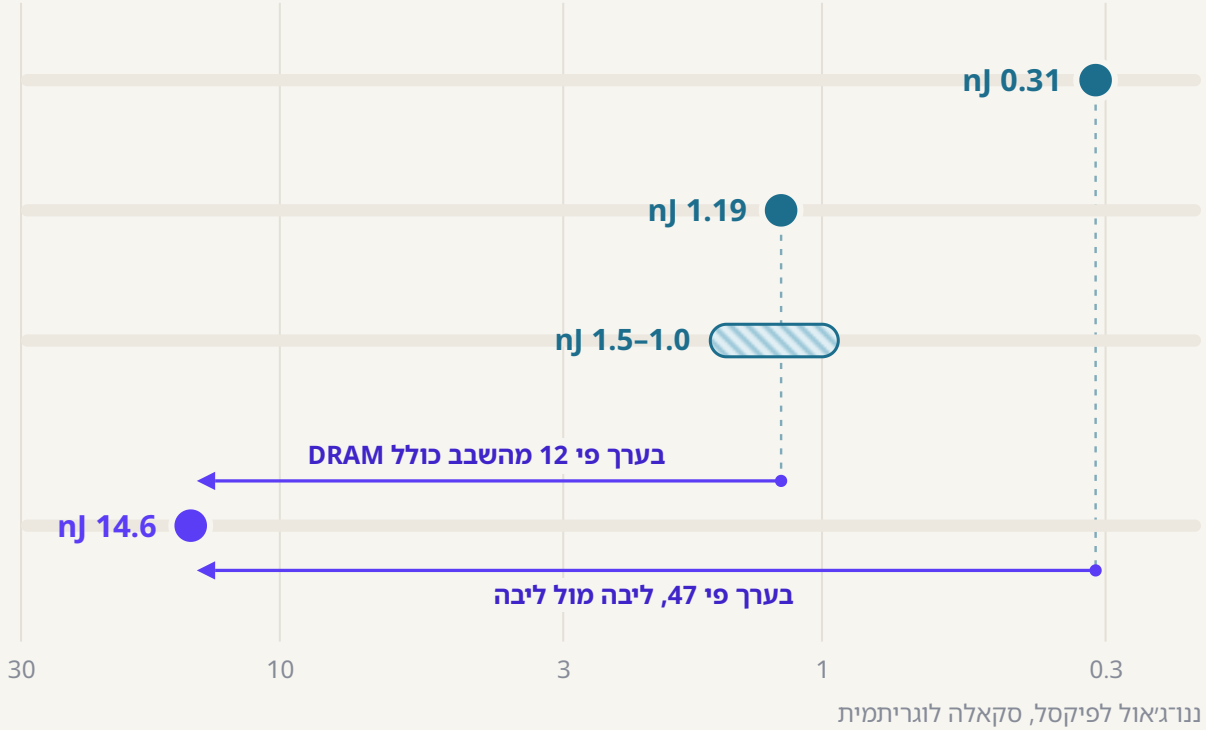
הודעה מס׳ 49 לשנת 2026, בתוקף מ־18 בנובמבר, מעדכנת את פיקוח הייצוא על רחפנים; אחד הסעיפים משתמש בדחיסת תמונה סמנטית כסף.

ארצות הברית

ה־FCC הוסיפה רחפנים מתוצרת חוץ ואת רכיביהם הקריטיים ל־Covered List שלה ב־22 בדצמבר 2025.

נישה מוצקה לחברה מתמחה עם יכולת ציות לפיקוח ייצוא, לא שוק רחב.

סדר גודל יותר אנרגיה לפיקסל



שבב מפענח HEVC, ליבה
40 nm, 4K30, נמדד

שבב מפענח HEVC, כולל DRAM
40 nm, 4K30, ליבה שנמדדה ו־DRAM מדומה

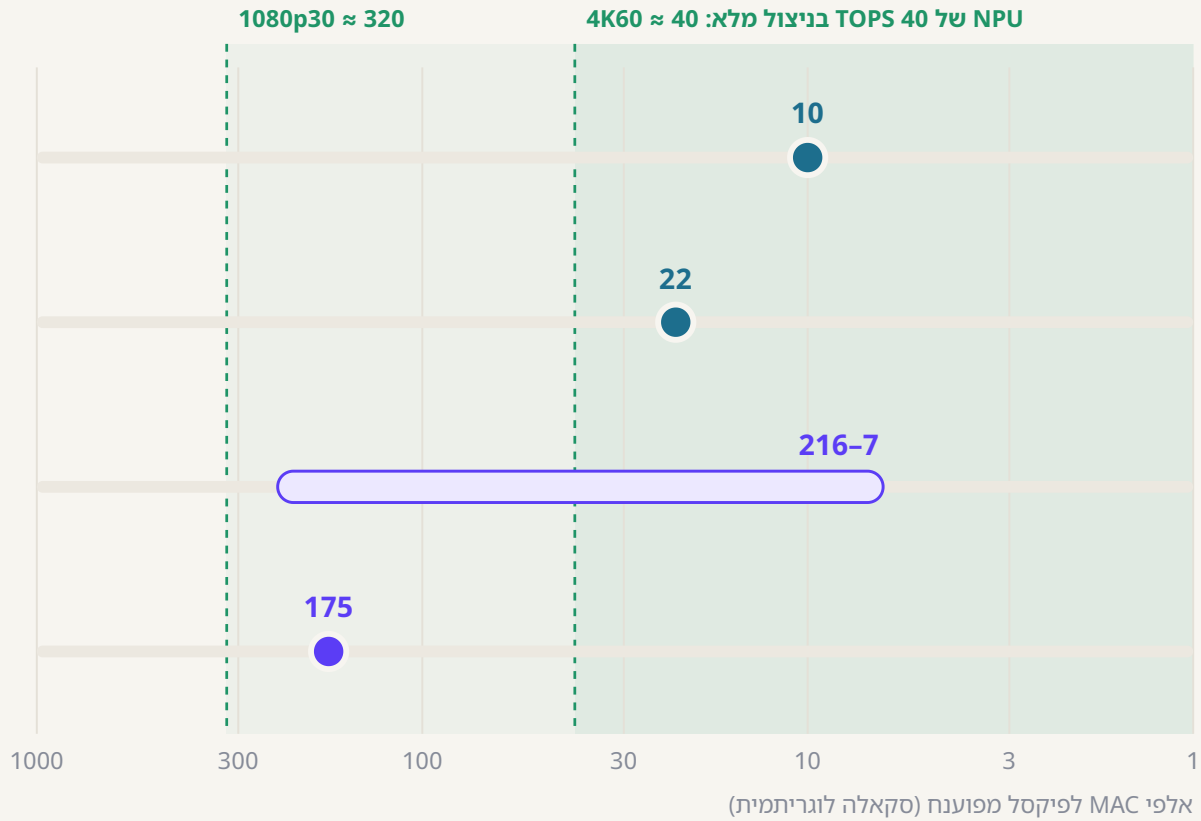
מפענח VVC
28 nm, 4K60, תחזית

מפענח וידאו נוירוני (NVCA)
28 nm, 1080p25, סינתזה בלבד; בלי משקלים ו־DRAM

ב־1080p30 זה כ־74 mW לעומת 908 mW.

ב־1080p30: כ־74 mW לשבב ה־HEVC כולל DRAM, לעומת כ־900 mW ל־NVCA. תהליך ייצור ישן יותר לשבב ה־HEVC, ומשקלים ו־DRAM שלא נספרו ב־NVCA: ההשוואה מחמיאה לתכנון הנוירוני.

מה מפענח נוירוני דורש מ־NPU



מסנן לולאה נוירוני, סיבוכיות נמוכה מאוד
ניסויי JVET · חיסכון של כ־6% בסיביות
1.2 TOPS ב־10 · 1080p30 ב־4K60

מסנן לולאה נוירוני, סיבוכיות נמוכה
ניסויי JVET · כ־8%
2.7 TOPS ב־22 · 1080p30 ב־4K60

מפענחי JPEG AI, מהבסיסי עד הגבוה
12.5–27.9% לעומת VVC תוך־תמונתי

DCVC-RT, מפענח וידאו נוירוני שלם
21% לעומת מקודד הייחוס של VVC, בתנאי המאמר
22 TOPS ב־174 · 1080p30 ב־4K60

DCVC-RT צריך כ־175 אלף MAC לפיקסל: כ־22 TOPS ב־1080p30 ו־174 ב־4K60, בניצול מלא. **אומדן** NPU של 40 TOPS בניצול מלא מאפשר כ־40 אלף ב־4K60.

תעבורת זיכרון וזהות ברמת הסיבית הן המגבלות הקשות

ה־DRAM שולט

זיכרון לפני MAC

בשבב ה־HEVC של MIT, ה־DRAM אחראי לכשלושה רבעים מאנרגיית המערכת: 0.88 מתוך 1.19 nJ לפיקסל. מפענחים נוירוניים מעבירים מפות מאפיינים גדולות בין פריימים, ומחברי DCVC-RT מצאו שקלט/פלט הזיכרון, ולא החישוב, הגביל את המהירות.

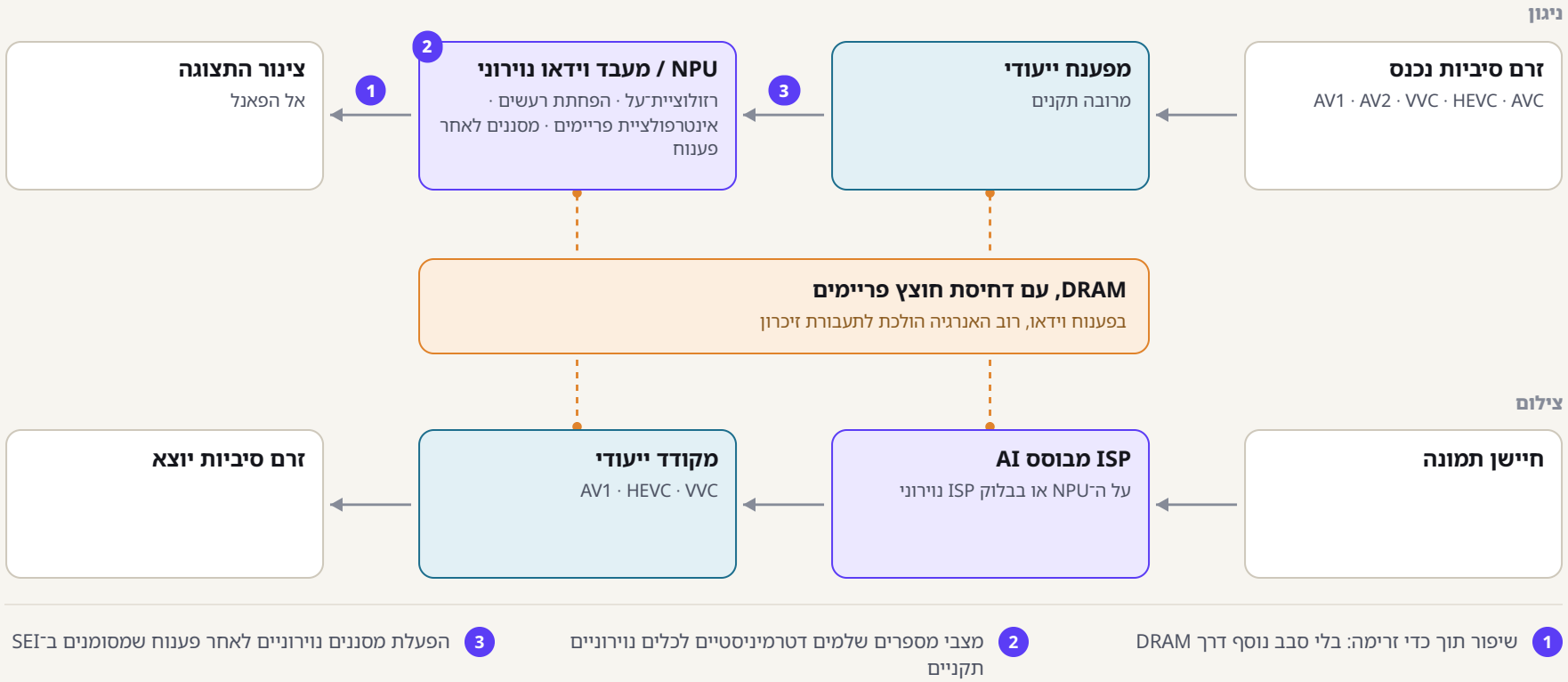
דיוק מוחלט

כל מפענח חייב לחשב את אותן סיביות

פענוח האנטרופיה נכשל אם שני מכשירים מחשבים הסתברות באופן שונה. הפתרונות: מודלים במספרים שלמים (DCVC-RT), hyperprior עם קנה מידה משותף (MLVC), הסקה במספרים שלמים של 16 סיביות (JVET של SADL), נקודות התאמה (JPEG AI).

לצוות שבבים זו שוב המשמעת של התמרות המספרים השלמים ב־HEVC: התקן חייב להגדיר את האריתמטיקה במדויק.

החלוקה שבשטח: קודק ייעודי ולצדו NPU



Canghai של Tencent: יותר מ־100,000 שבבים מהדור הראשון בפריסה; הדור השני מוסיף H.266 ורזולוציית־על על השבב. Chips&Media ו־Allegro DVT מוכרות IP של קודקים ו־IP של וידאו נוירוני זה לצד זה.

מה הייתי בונה, לפי תפקיד

צוותי SoC ושבבי טלוויזיה

- מפענחים ייעודיים לכל תקן שחייבים לנגן, AV2 כבר עכשיו
- NPU בגודל שמתאים לשיפור ברזולוציית התצוגה, בלי סבב נוסף דרך DRAM
- SEI של מסננים נוירוניים לאחר פענוח, ומצבי מספרים שלמים דטרמיניסטיים
- H.267 כתחילת תכנון ב־2029 או אחר כך

ספקי IP

- מפענחי AV2 ו־VVC עכשיו
- מאיצים למסננים לאחר פענוח ודחיסת חוצץ פריימים
- כלי בדיקת התאמה לעיבוד נוירוני

פלטפורמות וידאו

- עיבוד מקדים נוירוני במקודד מחזיר את ההשקעה ראשון
- רזולוציית־על בצד הלקוח היכן שהמכשירים מאפשרים
- קודקים נוירוניים מקצה לקצה רק במעגלים סגורים שבשליטתכם

היכן כלים נוירוניים נעשים ניתנים לרישוי

תקן	תוכן נוירוני	מצב, נובמבר 2026
H.267 (שם עבודה)	פתוח: כלים קלאסיים (ECM) וכלים נוירוניים מועמדים שניהם	הערכת המענים בינואר 2027; גרסה 1 בסוף 2029
VVC ו־HEVC, עם VSEI	מסנן נוירוני לאחר פענוח, מתואר ב־SEI; אופציונלי למפענחים	בתוקף; הורחב באוקטובר 2025
AV2	קלאסי, ללא תמלוגים	גרסה 1.0.0 ב־28 במאי 2026; סטרימינג של YouTube באתר מתוכנן לתחילת 2027
JPEG AI	קודק תמונה נלמד מקצה לקצה	חלק 1 פורסם ב־2025; לא נמצאה תמיכה כברירת מחדל במוצרים
IEEE 1857.11	קידוד תמונה נוירוני שתוכנן ל־NPU	פורסם בדצמבר 2024
Video Coding for Machines	קידוד לניתוח ממוחשב ולא לצפייה	טיוטת תקן בינלאומי בתחילת 2026
הרחבות ה־ML של Opus	אופציונליות ותואמות לאחור	ב־libopus 1.5 ו־1.6

645

משפחות פטנטים בעבודה שמעבר ל־VVC בידי ByteDance (GreyB, 2025)

0.50 · 1.50–1.00 דולר

תעריף המאגר של VVC לטלפון ולטלוויזיית 4K, אזור 1

0 דולר

תמלוגים על AV1 ו־AV2 מחברות AOMedia

הכסף הולך בעקבות התקנים והפלטפורמות

מרוויחים כבר עכשיו

פלטפורמות

שומרות את החיסכון

מקטין הרזולוציה הנוירוני של Netflix; רזולוציית־על ב־YouTube וב־Microsoft; Meta פתחה את הקוד של MLVC ברישיון MIT

מרוויחים כבר עכשיו

יצרני מכשירים ושבבים

מוכרים בידול

מעבדי ISP למצלמות; מעבדי הגדלת רזולוציה לטלוויזיות; שבבי הקודק Canghai שפיתחה Tencent בעצמה

מרוויחים כבר עכשיו

ספקי IP

מוכרים רישיון לכל תקן חדש

לקוחות למפענחי AV2 אצל Chips&Media, Allegro DVT ו־VeriSilicon בתוך שלושה שבועות מהמפרט

מרוויחים כבר עכשיו

ספקי GPU

בונים תלות בפלטפורמה

DLSS 4 ביותר מ־250 משחקים ואפליקציות; SDK לטקסטורות נוירוניות בבטא

לפי מכשיר

בעלי פטנטים ומאגרים

גובים לפי מכשיר

VVC: 0.50 דולר לטלפון, 1.00–1.50 דולר לטלוויזיית 4K; כלי H.267 אחרי 2029

אקזיט ברכישה

סטארט־אפים של קודקים נוירוניים

נרכשו

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

שלושת הסטארט־אפים המוכרים ביותר בתחום הקודקים הנוירוניים, WaveOne, iSIZE ו־Deep Render, נקנו על ידי Apple, Sony ו־InterDigital.

שש הזדמנויות, והסיכון בכל אחת

הזדמנות	מי קונה	למה עכשיו	הסיכון העיקרי
עיבוד מקדים נוירוני במקודד	פלטפורמות סטרימינג ווידאו של משתמשים, גופי שידור	בלי שינוי במכשירים; הוכח ב־Netflix	פלטפורמות גדולות בונות זאת בעצמן
IP וכיוונון לשיפור בצד התצוגה	יצרני שבבים לטלוויזיות, לממירים ולטלפונים	ה־NPU כבר נמצא במכשירים	הפיכה למוצר מדף; צוותים פנימיים
קודקים נוירוניים במעגל סגור	ספקי תקשורת ארגונית, תעשייה ונוכחות מרחוק	MLVC מוכיח שזה עובד, והקוד שלו פתוח	חייבים לשלוט בשני הקצוות; פיצול בין NPU שונים
סיליקון או IP של קודק ו־NPU	יצרני SoC	מחזור AV2 מתחיל	שחקנים ותיקים וחזקים
הערכה, בדיקות התאמה והסקה דטרמיניסטית	כל מי שמשווק כלים נוירוניים	זהות ברמת הסיבית היא דרישה מחייבת	שוק קטן, אלא אם זה חלק מחבילה
תרומות לתקינה של H.267	מעניקי רישיונות, פלטפורמות גדולות	המענים יוערכו בינואר 2027	ארוך, יקר, לא ודאי

קחו טענה בעירבון מוגבל, עד שהיא מציינת ארבעה דברים

טענה	מה הראיות אומרות
"קודקים נוירוניים כבר מנצחים את VVC."	במעבדה. DCVC-RT חוסך 21% לעומת מקודד הייחוס של VVC; המפענח שלו עולה כ־175 אלף MAC לפיקסל.
"קודקי AI יחליפו את H.26x ואת AV1 בתוך כמה שנים."	לא בניגון: גרסה 1 של H.267 בסוף 2029, שבבים בסביבות 2031 **אומדן**. הקודק שמגיע עכשיו, AV2, הוא קלאסי.
"MLVC חוסך 88% מרוחב הפס."	לעומת H.264 ב־360p. לעומת HEVC בחומרה: 75.5% ב־360p ו־65.4% ב־540p, על מערך הבדיקה של Microsoft עצמה.
"טקסטורות נוירוניות חוסכות 85–96% מזיכרון הווידאו."	בסצנות הדגמה. המשחק שיצא לשוק השיג כ־30%, ורק בנכסים נבחרים.
"JPEG AI הוא תקן ISO, ולכן מצלמות ודפדפנים יאמצו אותו."	לא נמצאה תמיכה כברירת מחדל במוצרים. האינטרנט אימץ את JPEG XL, פורמט קלאסי.

בקשו את נקודת הייחוס, המדד, הרזולוציה ועלות הפענוח. טענה שאינה יכולה לציין את כל הארבעה היא תוצאת מחקר, לא מוצר.

מה כדאי לעקוב אחריו, ושאלות פתוחות

ציוני דרך

ינואר 2027	הערכת המענים ל־H.267: כמה מהם נשענים על כלים נוירוניים, ובאיזו עלות לפיקסל
תחילת 2027	הסטרימינג הראשון של AV2 באתר של YouTube; מפענחי AV2 ב־SoC שהוכרזו
המחצית השנייה של 2027	מודל הבדיקה הראשון של H.267
2027	MLVC מעבר לשיחות עמית־לעמית ב־Teams
2028–2027	מוצר ראשון שמשתמש ב־JPEG AI כברירת מחדל
סוף 2029	גרסה 1 של H.267 פותחת את מחזור הסיליקון

שאלות פתוחות

01 דיוק מוחלט בקנה מידה רחב: פענוח במספרים שלמים, זהה ברמת הסיבית, על NPU של ספקים שונים, ובאיזה אובדן דחיסה?

02 זיכרון לפני MAC: תעבורת ה־DRAM לפיקסל של עיבוד נוירוני לאחר הפענוח

03 העלות לפיקסל שבה כלי נוירוני מפסיק להצדיק את עצמו ב־H.267

04 אילו מעגלים סגורים מעבר לשיחות יאמצו ראשונים קודקים נוירוניים מקצה לקצה

05 איך מודדים נאמנות כשמפענח מייצר פרטים

אלה תחומים שנוכל לחקור יחד, דרך חילופי ידע טכני ושיתופי פעולה טכניים.

i@tanshuai.com · tanshuai.com