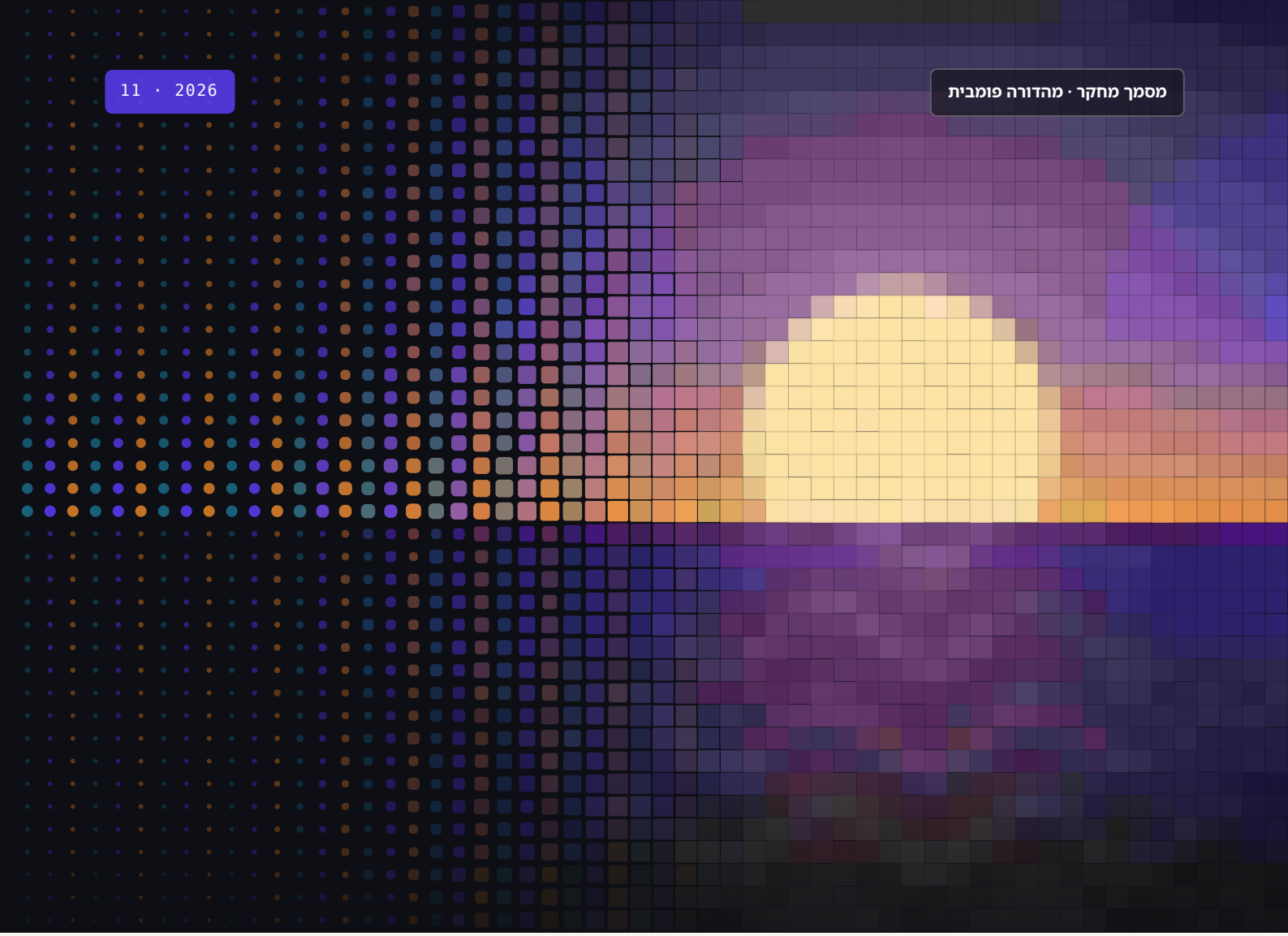


מסמך מחקר · מהדורה פומבית

11 · 2026

דחיסה נלמדת · ניגון × סיליקון

קידוד נוירוני ב־2026

בשטח הוא פועל סביב הקודק, לא במקומו

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

המסקנה

רשתות נוירונים כבר פועלות בשני הקצוות של הקודקים שנושאים את הווידאו של העולם. קודקים נוירוניים מקצה לקצה נמצאים בשימוש רק במעגלים סגורים, והמסכים הנפוצים ממתינים ל־H.267.

תאן שואי

טכנולוג ומייסד · חוקר ומפתח בינה מלאכותית

tanshuai.com · i@tanshuai.com

הנתונים נכונים ל־12 בנובמבר 2026

הסכומים בדולר ארה״ב

תוכן העניינים

תוכן העניינים



◆ **תקציר מנהלים** — כבר בשימוש רחב, אבל סביב הקודק — 04

01 **היקף ושיטה** — מה פירוש "קידוד נוירוני" כאן, ואיך בדקתי — 07

02 **ניגון וסטרימינג** — ניגון וסטרימינג: נוירוני בשני הקצוות, תקני באמצע — 08

03 **צילום ותמונות** — צילום ותמונות: נוירוני בתוך המצלמה, קלאסי באינטרנט — 11

04 **שיחות ופגישות** — שיחות ופגישות: היכן קודקים נוירוניים מקצה לקצה באמת נמצאים בשימוש — 12

05 **בינה מלאכותית יוצרת** — בינה מלאכותית יוצרת: הפריסה הגדולה ביותר, שאף אחד לא קורא לה קודק — 14

06 **משחקים וגרפיקה** — משחקים וגרפיקה: שחזור נוירוני כבר נפוץ, נכסים נוירוניים רק מתחילים — 15

07 **ערוצים ייעודיים** — ערוצים ייעודיים: לוויינים, רחפנים ורובוטים — 16

08 **סיליקון** — המבט מהסיליקון: אנרגיה, חישוב, זיכרון ודיוק מוחלט — 17

09 **תקנים ופטנטים** — תקנים ופטנטים: היכן כלים נוירוניים נעשים ניתנים לרישוי — 21

10 **מי מרוויח** — מי מרוויח, והיכן חברה יכולה לפעול — 22

11 **טענות וראיות** — טענות שתשמעו, ומה הראיות אומרות — 24

12 **ציוני דרך** — מה כדאי לעקוב אחריו, 2027 עד 2030 — 25

13 **שאלות פתוחות** — 26

◆ **על המחבר** — תאן שואי — 27

◆ **מקורות** — 28

איורים

איורים



איור 01	היכן קידוד נוירוני נמצא בשימוש ב־2026	06
איור 02	שלוש דרכים שבהן רשת נוירונים נכנסת לקודק	07
איור 03	הדרך מתקן גמור למסכים הנפוצים אורכת רוב עשור	10
איור 04	בשיחות, קודק נוירוני עדיף על HEVC בחומרה ב־65–76%	13
איור 05	כשחברה אחת שולטת בכל השכבות, שחזור נוירוני מתפשט מהר	15
איור 06	אנרגיה לפיקסל מפוענח: מפענחים ייעודיים מובילים בסדר גודל	18
איור 07	כלי נוירוני עולה עשרות אלפי MAC לפיקסל; מפענח נוירוני שלם עולה מאות אלפים	19
איור 08	החלוקה שבשטח: חומרה ייעודית לזרם הסיביות, מנוע נוירוני לפיקסלים	20
איור 09	מי מרוויח מקידוד נוירוני ב־2026	22

תגיות בטקסט

- **אומדן** — חישוב או הסקה של המחבר מתוך נתונים מצוטטים
- **טענת החברה** — פורסם על ידי החברה שמוכרת את המוצר, ולא אומת באופן בלתי תלוי
- **ניתוח** — הפרשנות של המחבר לראיות

זהו תרגום של המהדורה האנגלית; בכל מקרה של אי־התאמה, המהדורה האנגלית היא הקובעת.

הנתונים נכונים ל־12 בנובמבר 2026 · הסכומים בדולר ארה״ב

תקציר מנהלים

כבר בשימוש רחב, אבל סביב הקודק

המסקנה

קידוד נוירוני כבר נמצא ב־2026 בשימוש רחב, אבל הוא פועל סביב הקודקים התקניים, לא במקומם.

כ־75%
מתעבורת הנתונים הסלולרית היא וידאו, סוף 2025 (Ericsson)

1.26 מיליארד
סמארטפונים שווקו ב־2025, וכל אחד מהם מפענח וידאו בחומרה ייעודית

פי 12 בערך
מהאנרגיה לפיקסל של שבב מפענח HEVC שנמדד כולל DRAM, בתכנון השבב היחיד שפורסם של מפענח נוירוני

כ־22 TOPS
של חישוב רציף למפענח נוירוני מסוג DCVC-RT ב־1080p30 אומדן

75.5%
פחות קצב סיביות מ־HEVC בחומרה עבור MLVC ב־360p בשיחות, על מערך הבדיקה של Microsoft עצמה טענת החברה

סוף 2029
גרסה 1 של H.267, התקן הראשון שצפוי להביא כלים נוירוניים לכל מסך

היכן הוא בשימוש

רשתות נוירונים רצות כיום בשני הקצוות של הקודקים שנושאים כמעט את כל הווידאו בעולם. בדרך פנימה הן מעצבות את התמונה: מעבדי אותות תמונה מבוססי AI במצלמות של טלפונים, ומקטין רזולוציה נוירוני בצינור הקידוד של Netflix. בדרך החוצה הן מתקנות אותה: Super Resolution של YouTube בטלוויזיות, רזולוציית־על ב־Meta, ומעבדי טלוויזיה שמפורסמים כבעלי מאות רשתות נוירונים. ביניהם, זרם הסיביות הוא עדיין AV1, HEVC, AVC או VVC, שמפוענח בחומרה ייעודית. אף שירות סטרימינג אינו שולח פורמט וידאו נוירוני לטלוויזיות או לטלפונים; היכן שמפענח וידאו נוירוני רץ היום על מכשיר, הוא רץ בתוך האפליקציה של חברה אחת.

מעגלים סגורים

קודקים נוירוניים מקצה לקצה, שבהם גם המקודד וגם המפענח נלמדים, נמצאים בשימוש רק במעגלים סגורים, שבהם חברה אחת שולטת בשני הקצוות ויכולה לחזור לקודק תקני. Microsoft מריצה קודק שמע מבוסס AI בכל השיחות בין שני משתתפים ב־Teams וב־Skype מאז 2021, וביולי 2026 פתחה את הקוד של MLVC, קודק וידאו נוירוני שנמצא עכשיו בפריסה בשיחות עמית־לעמית ב־Teams: ב־360p הוא צריך קצב סיביות נמוך בכ־75% מזה של HEVC בחומרה לאותו ציון איכות. כל מודל ליצירת דיבור ווידאו רץ על קודק נלמד, מטוקנייזרים של דיבור ב־12.5 Hz ועד VAE לווידאו עם יחס דחיסה של 1:192. דחיסת טקסטורות נוירונית כבר יצאה לשוק במשחק AAA אחד. אלה פריסות אמיתיות, ויש להן תנאי משותף אחד: אף אחד אחר לא צריך לפענח את הזרם.

תזמון

ניגון המוני של זרמי סיביות נוירוניים ממתין לתקן ולמחזור סיליקון. התקן הראשון שצפוי לשאת כלים נוירוניים, יורשו של VVC שאכנה H.267, פרסם קול קורא להצעות ביולי 2026, יעריך את המענים בינואר 2027 ומכוון לגרסה ראשונה בסוף 2029. ל־AV1 נדרשו כמעט שמונה שנים מהמפרט הגמור ועד 30% מהצפייה ב־Netflix. לפי הדפוס הזה, שבבי H.267 הראשונים יגיעו בסביבות 2031, והשימוש הרחב יבוא

באמצע שנות ה־2030 **אומדן**. בינתיים, הקודק הבא שיגיע למסכים הוא AV2, שהושלם במאי 2026 בלי מפענח נוירוני, ושלושה ספקי IP הכריזו על לקוחות למפענחי AV2 בתוך שלושה שבועות.

סיליקון

הסיבה נמצאת בסיליקון. תכנון השבב היחיד שפורסם למפענח וידאו נוירוני צורך 14.6 nJ לפיקסל, בערך פי 12 משבב מפענח HEVC שנמדד כשסופרים גם את ה־DRAM שלו, ובערך פי 47 מהליבה שלו לבדה. DCVC-RT, קודק המחקר של Microsoft לזמן אמת, צריך, לפי הספירה שלי מטבלת המודולים שלו עצמו, כ־175 אלף פעולות הכפלה־וצבירה לכל פיקסל מפוענח, כלומר כ־22 TOPS של חישוב רציף עבור 1080p30 **אומדן**: בערך מחצית מ־40 ה־TOPS שמגדירים מחשב AI, שמוקדשת לזרם וידאו אחד. מפענח נוירוני גם חייב להפיק תוצאות זהות ברמת הסיבית בכל מכשיר, אחרת פענוח האנטרופיה שלו מתפרק. זו הסיבה ש־MLVC מוותר על דחיסה כדי להתפענח באופן זהה על NPU שונים, ורץ ב־540p ולא ב־4K.

עסקים

מבחינה עסקית, הכסף הולך בעקבות התקנים והפלטפורמות, לא בעקבות קודקים נוירוניים עצמאיים. הפלטפורמות שומרות לעצמן את החיסכון מעיבוד נוירוני לפני הקידוד ואחרי הפענוח. ספקי IP מוכרים רישיונות למפענחים ייעודיים לכל תקן חדש, וגם NPU לעיבוד תמונה. מאגרי פטנטים גובים לפי מכשיר: 0.50 דולר לטלפון ו־1.00–1.50 דולר לטלוויזיית 4K עבור VVC, בארה״ב ובשאר השווקים של אזור 1. שלושת הסטארט־אפים המוכרים ביותר של קודקים נוירוניים, iSIZE ,WaveOne ו־Deep Render, נקנו על ידי Apple, Sony ו־InterDigital. ההזדמנויות הפתוחות נמצאות בעיבוד נוירוני לפני הקידוד ואחרי הפענוח, במוצרים במעגל סגור, בסיליקון שמשלב קודק ייעודי עם NPU, ובעבודת ההערכה והתקינה שביניהם.

שאלה	הכרעה	מספרי מפתח (המקורות בפרקים)
1 **היכן קידוד נוירוני נמצא היום בשימוש בהיקף רחב?**	סביב הקודק: ISP במצלמות, הקטנת רזולוציה בצד המקודד, רזולוציית־על בניגון. מקצה לקצה רק במעגלים סגורים: שיחות, נכסי משחק, טוקנייזרים של בינה מלאכותית יוצרת	וידאו הוא כ־75% מהנתונים הסלולריים; 1.26 מיליארד טלפונים וכ־200 מיליון טלוויזיות שווקו ב־2025, וכולם מפענחים בחומרה ייעודית; DLSS 4 ביותר מ־250 משחקים ואפליקציות
2 **האם קודקים נוירוניים יחליפו את AV1, HEVC ו־VVC בניגון? מתי?**	לא לפני תקן ומחזור מכשירים. H.267 הוא הנשא הריאלי הראשון, וסביר שישלב כלים קלאסיים ונוירוניים	קול קורא להצעות ל־H.267 ביולי 2026, הערכה בינואר 2027, גרסה 1 בסוף 2029; שימוש רחב באמצע שנות ה־2030 **אומדן**
3 **כמה זה עולה בסיליקון?**	סדר גודל יותר אנרגיה לפיקסל מפענוח ייעודי היום, ובנוסף תעבורת זיכרון ודרישה לזהות ברמת הסיבית	14.6 nJ לפיקסל (נוירוני) לעומת 1.19 (HEVC, כולל DRAM); כ־175 kMAC לפיקסל מפוענח ≈ 22 TOPS ב־1080p30 **אומדן**
4 **מה התקנים והפטנטים מאפשרים?**	כבר אפשר לסמן מסננים נוירוניים לאחר פענוח בזרמי VVC; JPEG AI הוא תקן ISO; H.267 יכריע אילו כלים נוירוניים יהפכו לניתנים לרישוי	ECM (קלאסי) כ־28% לעומת VVC; מסנני לולאה נוירוניים 6–14% ב־10–22 kMAC לפיקסל ומעלה; ל־ByteDance יש 645 משפחות פטנטים בעבודה שמעבר ל־VVC
5 **מי מרוויח, והיכן חברה יכולה לפעול?**	פלטפורמות, ספקי IP, בעלי פטנטים. סטארט־אפים שעסקו אך ורק בקודקים נוירוניים נרכשו. לפעול בעיבוד לפני הקידוד ואחרי הפענוח, במעגלים סגורים, בסיליקון של קודק ו־NPU, בהערכה	מאגר VVC: 0.50 דולר לטלפון; IP של מפענחי AV2 נמכר ברישיון בתוך שבועות מפרסום המפרט; שלוש רכישות, בתנאים שלא פורסמו
6 **אילו טענות כדאי לקחת בעירבון מוגבל?**	שיפורים שמצוטטים לעומת קודקים ישנים, ברזולוציה נמוכה, בלי עלות הפענוח, או מהדגמות במקום ממוצרים שיצאו לשוק	MLVC: 88% פחות מ־H.264 אבל 65–76% פחות מ־HEVC בחומרה; טקסטורות: חיסכון של 85% ב־VRAM בהדגמה, כ־30% במשחק שיצא לשוק

איור 01 היכן קידוד נוירוני נמצא בשימוש ב־2026

רשתות נוירונים כבר רצות בשני הקצוות של הקודקים הנפוצים. קודקים נוירוניים מקצה לקצה נמצאים בשימוש רק היכן שחברה אחת שולטת גם במקודד וגם במפענח.

	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	—

מקור: מסמכי חברות, תקנים ומפתחים שמצוטטים בדוח; המצב נכון ל־12 בנובמבר 2026. השורות מסודרות לפי מספר האנשים שכל תחום מגיע אליהם.

היקף ושיטה

01 מה פירוש "קידוד נוירוני" כאן, ואיך בדקתי

שלוש דרכים שבהן רשת נוירונים נכנסת לקודק

"קידוד נוירוני" מכסה שלושה דברים שונים, ורוב הוויכוחים עליו נובעים מערבוב ביניהם.

01 **סביב הקודק.** רשת מעבדת פיקסלים לפני הקידוד או אחרי הפענוח, וזרם הסיביות נשאר תקני. דוגמאות: מעבד אותות תמונה (ISP) מבוסס AI במצלמה, מקטין הרזולוציה הנוירוני של Netflix לפני המקודד שלה, רזולוציית־על בטלוויזיה. כל מכשיר יכול לנגן את הזרם; הרשת משנה רק את מה שנכנס או את מה שיוצא.

02 **בתוך תקן.** רשת הופכת לכלי בקודק תקני, למשל מסנן לולאה נוירוני או מנבא תוך־תמונתי שכל מפענח חייב להריץ, ולכן יש להגדיר אותה במדויק, בדרך כלל באריתמטיקה של מספרים שלמים. צורה קלה יותר היא מסנן נוירוני לאחר פענוח שזרם הסיביות מתאר במטא־נתונים, והמפענחים יכולים לבחור אם להפעיל אותו.

03 **במקום הקודק.** גם המקודד וגם המפענח נלמדים מקצה לקצה. זרם הסיביות הוא אוסף של ערכים לטנטיים מכומתים שרק הרשת התואמת יכולה לפענח. דוגמאות: MLVC של Microsoft ב־Teams, טוקנייזרים של דיבור במודלי קול, טקסטורות נוירוניות במשחק.

איור 02

שלוש דרכים שבהן רשת נוירונים נכנסת לקודק

רוב מה שנמצא בשימוש ב־2026 פועל סביב קודק תקני. כלים נוירוניים בתוך תקן נמצאים בשלבי הגדרה. קודקים נוירוניים מקצה לקצה מחליפים את הקודק רק היכן שחברה אחת שולטת בשני הקצוות.



מקור: סינתזה של המחבר מתוך המקורות בדוח זה; המצב נכון ל־12 בנובמבר 2026.

הסוג הראשון אינו צריך רשות מאיש. השני צריך תקן. השלישי צריך שליטה בשני הקצוות. זה הסדר שבו הם מגיעים לאנשים, והוא מסביר את רוב מה שיבוא בהמשך.

לפי סדר העולם האמיתי, לא לפי סדר היום של המחקר

מאמרי מחקר עוסקים בעיקר בסוג השלישי, כי שם נמצאים שיאי הדחיסה. הדוח הזה מסדר את התחומים לפי מספר האנשים שהם מגיעים אליהם, כי זה מה שקובע לאן הולכים הכסף וההנדסה: קודם ניגון וסטרימינג, אחר כך צילום, שיחות, בינה מלאכותית יוצרת, משחקים, ובסוף הערוצים הייעודיים (לוויין, רחפנים, רובוטים). וידאו הוא כ־75% מכל תעבורת הנתונים הסלולרית, לפי הספירה של Ericsson בסוף 2025. ב־2025 שיווקה התעשייה כ־1.26 מיליארד סמארטפונים (IDC) וכ־196–209 מיליון טלוויזיות (לפי אומדני TrendForce ו־Omdia). כל אחד מהם מפענח וידאו בחומרה ייעודית.

איך עבדתי

זהו מחקר שולחני, שנעשה כפי שהייתי בוחן שבב או תוכנית מוצר. קראתי מסמכי תקינה ודוחות ישיבות (JVET, MPEG, JPEG, AOMedia, IETF), בלוגים הנדסיים והודעות גרסה של חברות, מאמרים על שבבים, טבלאות תעריפים של מאגרי פטנטים, טקסטים של פיקוח ייצוא ואת העיתונות. לכל מספר מצוין התאריך שלו. היכן שגזרתי מספר בעצמי, כמו חישוב לפיקסל מתוך טבלאות של מאמר, אני אומר זאת ומסמן אותו **אומדן**. טענות שלא יכולתי לבדוק מול מקור ראשוני מסומנות **טענת החברה**. הסכומים בדולר ארה״ב. המצב נכון ל־12 בנובמבר 2026.

אני משווה בין דברים דומים היכן שאני יכול, ואומר זאת היכן שאינני יכול. לשיפורי דחיסה אין משמעות רבה בלי ארבעה דברים שמצורפים אליהם: נקודת הייחוס (שיפור לעומת מה), המדד (PSNR, VMAF או פאנל של צופים), התוכן והרזולוציה, וכמה עולה הפענוח. רוב המחלוקת על קודקים נוירוניים נעלמת ברגע שמציינים את ארבעתם.

פרק 02

02

ניגון וסטרימינג

ניגון וסטרימינג: נוירוני בשני הקצוות, תקני באמצע

מה שהצופים מקבלים הוא עדיין זרם סיביות תקני

שירותי הווידאו הגדולים ביותר השלימו מעבר קודק אחד ומתחילים את הבא, ואף אחד מהם אינו כרוך במפענח נוירוני. בדצמבר 2025 דיווחה Netflix ש־AV1 נשא כ־30% מהצפייה בה, והיה הקודק השני בשכיחותו אצלה אחרי AVC. AV1 הגיע לטלוויזיות רק כשמפענחי החומרה הגיעו אליהן: Netflix הוסיפה את AV1 להסמכת המכשירים שלה ב־2019, הזרימה אותו לטלוויזיות מסוף 2021, ואומרת ש־88% מהמכשירים עם מסך גדול שהוגשו להסמכה בין 2021 ל־2025 תמכו בו, כמעט כולם מאז 2023. YouTube אמרה שנכון ל־19 במרץ 2026, יותר מ־90% מהקטלוג שלה, משוקלל לפי זמן צפייה, היו זמינים ב־AV1.

AV2, יורשו של AV1 מבית Alliance for Open Media, הגיע לגרסה 1.0.0 ב־28 במאי 2026, עם דחיסה טובה בכ־30% מזו של AV1 לפי המדידות האובייקטיביות של AOMedia. YouTube מתכננת סטרימינג של AV2 באתר ״כבר בתחילת 2027״. הסיליקון הגיע מיד אחריו: Chips&Media השלימה מפענח חומרה ל־AV2 במאי, וביוני מכרה רישיון שלו לחברת ביג־טק צפון־אמריקאית שזהותה לא נמסרה; Allegro DVT מסרה את ה־IP של מפענח ה־AV2 שלה לשלושה לקוחות עד 16 ביוני; VeriSilicon שחררה IP וידאו עם תמיכה ב־AV2 לכמה לקוחות עד 10 ביוני. AV2 הוא קודק קלאסי: הרווחים שלו באים משיפורים בחלוקה, בחיזוי, בהתמרות ובמסנני הלולאה (אחד מהם מסנן פרטים מונחה ״מבוסס נתונים״), לא מרשת מקצה לקצה.

VVC, התקן הנוכחי של ITU/ISO, מצא את בית השידור הראשון שלו בברזיל, שמערכת ה־(TV 3.0) DTV+ שלה עברה לפעילות מסחרית ביוני 2026, בעיתוי שכוון למונדיאל. נכון ליולי 2026, אף טלוויזיה שנמכרה בברזיל לא כללה מקלט DTV+ מובנה;

הראשונות צפויות בין סוף 2026 לתחילת 2027.

במקודד: מקטין הרזולוציה הנוירוני של Netflix

הדוגמה הנקייה ביותר לקידוד נוירוני בניגון פועלת מאז 2022, ואינה משנה דבר במה שהמכשיר של הצופה צריך לעשות. Netflix מקודדת כל כותר בכמה רזולוציות, ושלב הקטנת הרזולוציה שמפיק 1080p ,720p ו־540p ממאסטר 4K השתמש באופן מסורתי במסנן קבוע כמו Lanczos. Netflix החליפה אותו ברשת נוירונים קטנה, בלוק עיבוד מקדים ואחריו בלוק שינוי גודל, שאומנה לשפר את האיכות הנראית אחרי מגדיל רזולוציה רגיל במכשיר. Netflix מתארת זאת כפתרון שנכנס במקום הקיים: אין שום שינוי אחר בצד הקידוד או בצד הלקוח, כך ש"מיליוני מכשירים שתומכים בסטרימינג של Netflix נהנים מכך אוטומטית". בבדיקות ההעדפה של Netflix, כ־77% מהצופים בחרו בפלט של מקטין הרזולוציה הנוירוני **טענת החברה**. כדי לשמור על עלות סבירה בהיקף של Netflix, "הספיקו רק כמה שכבות של רשת נוירונים".

זה הדפוס שכדאי לזכור. הרשת רצה פעם אחת לכל כותר במרכז הנתונים, שם החישוב זול ואפשר לתזמן אותו, והפלט שלה הוא זרם AV1 או HEVC רגיל.

במסך: רזולוציית־על באפליקציה ובטלוויזיה

בקצה השני, רשתות מתקנות את מה שהקודק והמקור איבדו. ב־29 באוקטובר 2025 הכריזה YouTube על Super Resolution לאפליקציית הטלוויזיה שלה: סרטונים שהועלו ברזולוציה נמוכה מ־1080p, מ־240p עד 720p, מוגדלים ל־HD בעזרת למידת מכונה, ו־4K הוא היעד המוצהר. הגרסאות המוגדלות מסומנות, הצופים יכולים לחזור למקור, והיוצרים יכולים לבחור שלא להשתתף. Meta תיארה באוקטובר 2025 איך היא מריצה רזולוציית־על לווידאו גם בשרתים שלה בזמן ההעלאה, על מעבדי x86 רגילים בשיתוף עם Intel, וגם אצל הלקוח בזמן הניגון, ומצאה ששתי הגישות משלימות זו את זו. יצרני הטלוויזיות עושים את אותו הדבר בסיליקון: Samsung מפרסמת שמעבד ה־8K הדגל שלה, NQ8 AI Gen3, כולל 768 רשתות נוירונים להגדלת רזולוציה.

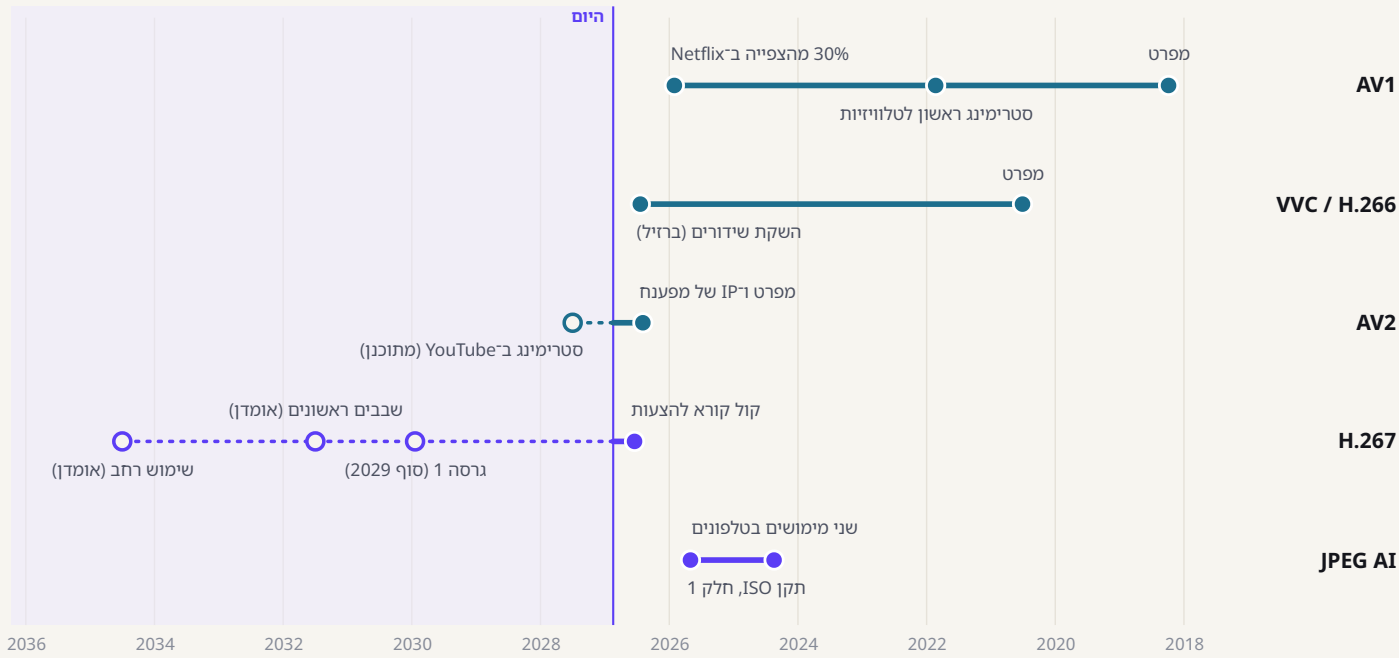
הכלכלה שונה לפי המקום שבו הרשת רצה. עיבוד בצד השרת עולה לפלטפורמה כוח חישוב על כל סרטון שהיא מטפלת בו. עיבוד בצד המכשיר אינו עולה לפלטפורמה דבר; הוא משתמש ב־NPU שהקונה כבר שילם עליו. זו הסיבה שהשיפורים הגדולים ביותר ב־2026 באים משניהם.

למה אף טלוויזיה אינה מפענחת זרם סיביות נוירוני

פורמט ניגון חייב להתפענח בכל מכשיר שינגן אותו אי פעם, והמכשירים האלה מחזיקים שנים. זה הופך את הדרך מתקן גמור למסכים הנפוצים לאיטית, אפילו עבור קודקים קלאסיים. המפרט של AV1 הושלם במרץ 2018; מפענח התוכנה dav1d הגיע בתוך חודשים; טלוויזיות קיבלו אותו בסטרימינג מסוף 2021; הוא הגיע לכ־30% מהצפייה ב־Netflix בדצמבר 2025, כמעט שמונה שנים אחרי המפרט. VVC הושלם ביולי 2020 והגיע לשירות השידור הראשון שלו ביוני 2026.

איור 03 **הדרך מתקן גמור למסכים הנפוצים אורכת רוב עשור**

ניגון מחייב מפענח חומרה במכשיר. ל־AV1 נדרשו כמעט שמונה שנים מהמפרט הגמור ועד 30% מהצפייה ב־Netflix; התקן הראשון שצפוי לשאת כלים נוירוניים מתוכנן לסוף 2029.



מקור: Netflix Tech Blog (דצמבר 2025); AOMedia; Allegro DVT; TV Technology; Fraunhofer IIS על DTV+ בברזיל; MPEG-155; ועדת ISO; JPEG. סימנים חלולים הם יעדים או אומדנים של המחבר.

זרם סיביות נוירוני מוסיף שתי בעיות שקודקים קלאסיים פתרו מזמן: אנרגיה לפיקסל ופענוח זהה ברמת הסיבית. שתיהן נדונות בפרק 8. יחד הן אומרות שאף יצרן טלוויזיות או טלפונים לא יכניס מפענח וידאו נוירוני לכל מכשיר לפני שתקן יגיד לו בדיוק מה לבנות.

מה ישנה את התמונה: H.267

התקן שיכול להביא כלים נוירוניים לכל מסך הוא יורשו של VVC, שמפותח במשותף על ידי ITU-T ו־ISO/IEC ב־Joint Video Experts Team (JVET). שמו עדיין אינו סופי; אכנה אותו H.267, כפי שעושה חלק גדול מהתעשייה.

באוקטובר 2025 הערי JVET את המענים לקול קורא לראיות: חמישה מענים מ־16 חברות ומעבדות מחקר, עם 20 תצורות קודק שנבדקו. הוועדה מצאה ראיות לדחיסה טובה יותר מ־VVC, והגדירה גם את הטכנולוגיה הקונבנציונלית המשופרת וגם את טכנולוגיית רשתות הנוירונים "מעניינות". הגישות הנוירוניות הראו שיפורים תפיסתיים ברורים, במיוחד בתוכן 8K HDR, ושיפורים קטנים יותר בתוכן משחקים. InterDigital, אחת המשיבות, הגישה תכנון אחד המבוסס על Enhanced Compression Model (ECM) הקונבנציונלי, ותכנון אחר שהוסיף כלי AI על גבי VVC.

הקול הקורא הרשמי להצעות פורסם בישיבת יולי 2026 בז׳נבה. מועד ההגשות העיקריות היה סוף אוקטובר 2026, חבילה משלימה צפויה באמצע דצמבר, והמענים יוערכו בישיבת JVET בבריסביין בינואר 2027. JVET מצפה למודל בדיקה ראשון במחצית השנייה של 2027 ולגרסה ראשונה של התקן עד סוף 2029, ומכנה את המאמץ "כנראה מאמץ גדול יותר מכל מה שנעשה אי פעם כדי להתניע תקן חדש".

המסלול הקלאסי מציב רף גבוה. ECM, שנבנה בעיבוד אותות קונבנציונלי, חוסך כ־28% לעומת VVC בתצורת גישה אקראית. ניסויי רשתות הנוירונים של JVET מדווחים על מסנני לולאה שחוסכים כ־6% ב־10 אלף פעולות הכפלה־וצבירה לפיקסל, 8% ב־22 אלף, ו־14% בנקודת עבודה בעלת סיבוכיות גבוהה. הקריאה שלי היא ש־H.267 יהיה ככל הנראה ליבה קלאסית עם כמה

כלים נוירוניים, היכן שהסיביות שנחסכות מצדיקות את עלות המפענח, וששההערכה בינואר 2027 תראה כמה כלים כאלה יהיו ניתוח.

כבר קיים גשר. זרמי VVC ו־HEVC יכולים לשאת מידע שיפור משלים (SEI) שמתאר מסנן נוירוני לאחר פענוח שצד המפענח יפעיל, והמהדורות החדשות של VVC ושל תקן ה־SEI שלו (VSEI), שהושלמו באוקטובר 2025, הרחיבו את ההודעות האלה. זרם יכול לציין איזו רשת להריץ אחרי הפענוח, בלי לחייב כל מפענח להריץ אותה.

מה זה אומר לעסקים בתחום הניגון

מכרו עכשיו את מה שעובד סביב הקודק: עיבוד מקדים בצד המקודד לשירותים, ובלוקים של עיבוד לאחר פענוח וכיוונון לשבבי טלוויזיות וממירים. אל תבנו עסק ניגון על זרם סיביות נוירוני שחייב להתפענח במכשירים של צד שלישי לפני שנות ה־2030. אם אתם מייצרים סיליקון, ההכנסות בטווח הקרוב הן ממפענחי AV2 ו־VVC; החלק הנוירוני הוא נתיב NPU לשיפור לצדם (פרק 8).

פרק 03

צילום ותמונות

03 צילום ותמונות: נוירוני בתוך המצלמה, קלאסי באינטרנט

צינור העיבוד של המצלמה כבר נוירוני

כל תמונה מטלפון דגל עדכני עוברת דרך רשתות נוירונים לפני שהיא נדחסת. Snapdragon 8 Gen 2 של Qualcomm, בטלפונים מ־2023, הציג "Cognitive ISP" שמריץ סגמנטציה סמנטית בזמן אמת במהלך צילום תמונות ווידאו, ומכוונן בנפרד פנים, שיער, שמיים ורקעים; Galaxy S23 Ultra של Samsung היה הטלפון הראשון שהשתמש בו במצלמת הסלפי. בינואר 2026 הציגו Chips&Media ו־Visionary.ai ב־CES את מה שהן מכנות מעבד ה־ISP המלא הראשון המבוסס על AI: צינור דימות מקצה לקצה שרץ כולו על NPU, ה־WAVE-N של Chips&Media, בנקודה צפה של 16 סיביות, בלי ISP ייעודי, ושאפשר לעדכן אותו מרחוק. "צינורות קיימים מצמידים בלוקים נוירוניים ל־ISP ייעודי. אנחנו מחליפים את ה־ISP הקונבנציונלי לחלוטין", אמר מנכ"ל Visionary.ai. אחרי התערוכה הוא הוצע ליצרנים לצורך הערכה.

אחר כך התוצאה נשמרת כ־JPEG או כ־HEIF, פורמטים קלאסיים שכל מכשיר יכול לפתוח. הדפוס זהה לזה של הניגון: נוירוני בתוך צינור העיבוד של חברה אחת, תקני בגבול.

JPEG AI: תקן נוירוני גמור, עדיין בלי מוצר

JPEG AI הוא תקן קידוד התמונה הנלמד הראשון של ISO/IEC ו־ITU-T (ISO/IEC 6048, ITU-T T.840). חלק 1, מערכת הקידוד המרכזית, פורסם ב־2025. ועדת JPEG מדווחת על שיפורי דחיסה של 12.5% עד 27.9% לעומת קידוד תוך־תמונתי ב־VVC, בטווח סיבוכיות מפענח של 7 עד 216 אלף פעולות הכפלה־וצבירה לפיקסל. בתכנון שלו יש שני מקודדים ושלושה מפענחים שאומנו במשותף, כך שמקלט יכול לממש מפענח אחד בלבד ועדיין לקרוא כל זרם. עד מאי 2024 דיווחה הוועדה על שני מימושים בסמארטפונים. במקביל, ה־IEEE פרסם בדצמבר 2024 את IEEE 1857.11-2024, תקן לקידוד תמונה מבוסס רשתות נוירונים שתוכנן ל־NPU.

נכון לנובמבר 2026 לא מצאתי מצלמת טלפון, אפליקציית תמונות או דפדפן שכותבים או קוראים JPEG AI כברירת מחדל.

מה שהאינטרנט אימץ במקום: JPEG XL

פורמט התמונה שאכן צבר תאוצה ב־2026 הוא קלאסי. ב־6 באוקטובר 2026 הפעילה Google פענוח JPEG XL כברירת מחדל ב־Chrome 155, עם מפענח חדש ובטוח־זיכרון שנכתב ב־Rust, ארבע שנים אחרי שהסירה את התמיכה הניסיונית. Safari מפענח JPEG XL מאז 2023, ו־Mozilla נעה באותו כיוון. הנימוקים של Google בעדו הם דחיסה טובה יותר מ־JPEG, דחיסה מחדש ללא אובדן של קובצי JPEG קיימים, HDR ופענוח הדרגתי.

הלקח הזה כללי. פורמט שחייב להיפתח בכל מקום מנצח עם מפענח מהיר ודטרמיניסטי, שרץ על כל CPU ואפשר להפוך אותו לבטוח־זיכרון. מפענח תמונה נוירוני צריך NPU או GPU ואריתמטיקה מדויקת בין מכשירים. קידוד נוירוני מנצח בתוך צינור העיבוד של המכשיר עצמו, ובמערכות שבהן אותה חברה מקודדת ומפענחת, כמו שירות תמונות בענן שלעולם אינו מוסר את הקובץ הדחוס לאף אחד אחר.

מה זה אומר לעסקים בתחום הדימות

הכסף נמצא בצינור הצילום: IP של ISP מבוסס AI, NPU בגודל שמתאים לדימות, ועבודת הכיוונון שיצרני המצלמות קונים. לאחסון ולהפצה, קודק תמונה נלמד הוא היום כלי של מערכת סגורה, לא פורמט שכדאי להמר עליו.



פרק 04

שיחות ופגישות

04 שיחות ופגישות: היכן קודקים נוירוניים מקצה לקצה באמת נמצאים בשימוש

השמע הגיע ראשון

Microsoft הציגה בפברואר 2021 את Satin, קודק דיבור מבוסס AI, ואמרה שהוא כבר משמש בכל השיחות בין שני משתתפים ב־Teams וב־Skype. Satin מספק דיבור בפס סופר־רחב, עד 16 kHz של רוחב פס שמע, ב־6 kbps, בעוד שקודק Silk הקודם של Skype היה צר־פס בקצב הזה. קצב הסיביות שהוא חוסך מושקע ביתירות, כך ששיחות שורדות טוב יותר אובדן מנות.

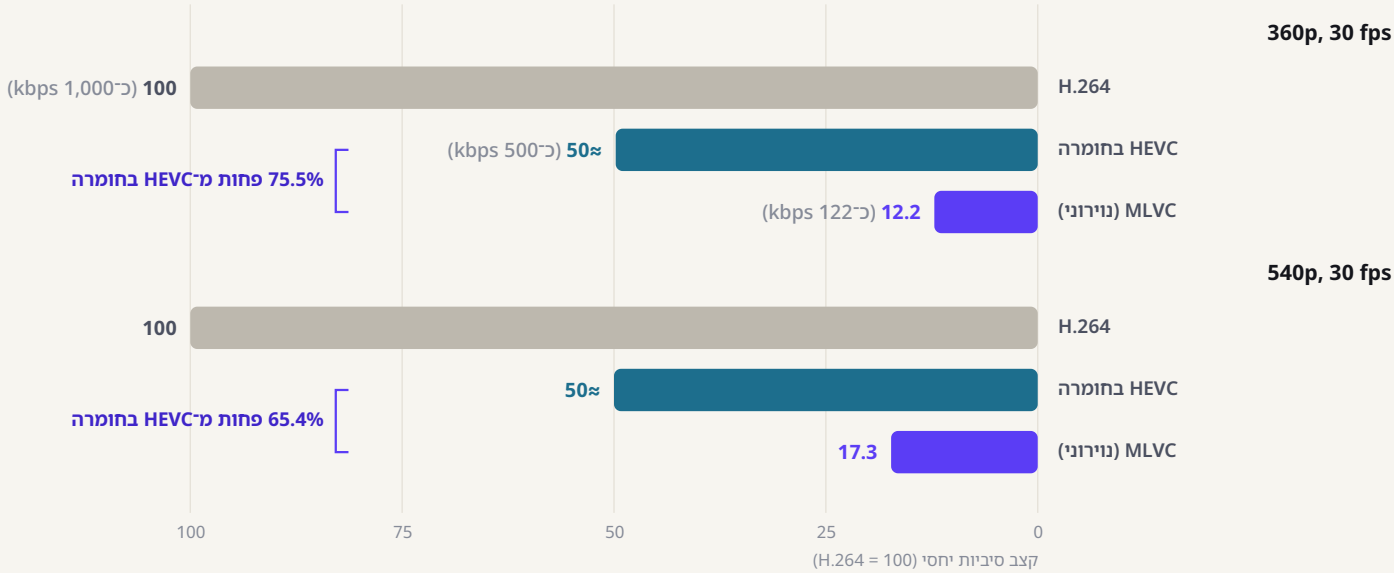
MLVC: קודק וידאו נוירוני בשיחות בסביבת ייצור

ב־24 ביולי 2026 שחררה Microsoft את MLVC, קודק הווידאו שלה מבוסס למידת המכונה, כקוד פתוח ברישיון MIT, יחד עם קוד המודל, המשקלים המאומנים, סקריפטי האימון וכלי ההמרה שמכוונים ל־NPU שונים. MLVC הוא גרסת המוצר של משפחת DCVC ש־Microsoft Research מפרסמת מאז 2021. Microsoft אומרת שהיא פורסת את MLVC ב־Teams לשיחות עמית־לעמית, עם טלמטריה חיה, בדיקות A/B והקודקים הקונבנציונליים כגיבוי.

MLVC רץ ב־540p וב־360p ב־30 פריימים לשנייה, והוא מהונדס לנצל לא יותר ממחצית מה־NPU ב־Neural Engine של Apple וב־NPU של Qualcomm ושל Intel, תוך חלוקת העבודה בין ה־CPU ל־NPU. בבדיקות סובייקטיביות לפי ITU-T P.910 על ה־Video Conferencing Dataset של Microsoft עצמה, הוא נזקק לקצב סיביות נמוך ב־87.8% מזה של H.264 וב־75.5% מזה של HEVC בחומרה ב־360p, ונמוך ב־82.7% וב־65.4% בהתאמה ב־540p. **טענת החברה** ב־360p מדובר בכ־122 kbps, היכן ש־H.264 צריך כ־1 Mbps.

איור 04 **בשיחות, קודק נוירוני עדיף על HEVC בחומרה ב־65–76%**

קצב הסיביות הנדרש לאותו ציון איכות על מערך הבדיקה של Microsoft לשיחות ועידה בווידאו, H.264 = 100. MLVC רץ על NPU של מחשבים ניידים וטלפונים, עד 540p30, ומנצל לכל היותר מחצית מה־NPU.



מקור: Microsoft, השקת MLVC (24 ביולי 2026): בדיקות סובייקטיביות לפי ITU-T P.910 על ה־Video Conferencing Dataset שלה. עמודות ה־HEVC נגזרו מהחיסכון המדווח. טענות החברה, לא מדידות בלתי תלויות.

שני דברים ממתנים את הכותרת. ה־88% הם לעומת H.264; לעומת HEVC בחומרה, שכבר נמצא באותם מחשבים ניידים, מדובר ב־65–76%, עדיין הרבה. ו־MLVC ויתר על דחיסה כדי שאפשר יהיה לפרוס אותו: ניתוח בלתי תלוי מציין ש־DCVC-RT שאומן לפי מדדים תפיסתיים משיג 81.9% באותו מדד שבו MLVC משיג 75.5%. המחיר של פענוח זהה על חומרה שאינה בבעלות Microsoft היה כשש נקודות.

למה השיחות הן הבית הטבעי

שיחה היא מעגל סגור. בשתי נקודות הקצה רצה אותה אפליקציה, שתיהן מתאמות את היכולות שלהן בתחילת השיחה, ויכולות לחזור לקודק תקני באמצעה. הרזולוציות צנועות, חיבורי הרשת לעיתים קרובות גרועים, והמחשבים הניידים בשני הקצוות מגיעים כיום עם NPU, כפי שטלפונים מגיעים כבר שנים. Microsoft מציבה 40 TOPS של NPU כרף למחשבי ה־AI שלה, וה־Neural Engine של M4 של Apple מדורג ב־38 TOPS. כל תנאי שחוסם קודק נוירוני בניגון נעדר כאן.

הדוגמאות הנגדיות: Opus ו־MLow

לא כל פלטפורמת שיחות עשתה את אותה בחירה, והסיבות מלמדות.

Opus, קודק השמע הפתוח שב־WebRTC וברוב אפליקציות השיחה, הוסיף למידה עמוקה בגרסה 1.5 (מרץ 2024), השימוש הראשון שלו ברשתות כדי לעבד או לייצר את האות עצמו. אבל הוא הוסיף אותן כאפשרויות: שיפור בצד המפענח (LACE ו־NoLACE) שאינו מחייב שינוי בזרם הסיביות, הסתרה נוירונית של אובדן מנות, ושיטת יתירות עמוקה (DRED); כולן כבויות כברירת מחדל, כדי לא להכביד על מכשירים ישנים. Opus 1.6 הוסיף הרחבת רוחב פס נוירונית ומודל DRED קטן פי שלושה, ונשאר תואם במלואו לתקן מ־2012. הגישה היא "בתוך תקן, אופציונלי ותואם לאחור".

Meta הלכה עוד יותר אחורה. ביוני 2024 השיקה את MLow, קודק דיבור לקצב סיביות נמוך שבנוי על עיבוד אותות קלאסי ולא על רשת, וכיום הוא פועל בכל השיחות ב־Instagram וב־Messenger ונמצא בפריסה ב־WhatsApp. הסיבות של Meta נעוצות במכשירים: יותר מ־20% מהשיחות שלה מתבצעות בטלפוני ARMv7, ועשרות מיליוני שיחות WhatsApp ביום מתבצעות

בטלפונים בני יותר מעשר שנים. קודקים נוירוניים, כולל EnCodec של Meta עצמה, רצים באופן אמין רק במכשירים יקרים, כתבה MLow. Meta מספק בערך פי שניים מהאיכות של Opus ב־6 kbps, ומשתמש ב־10% פחות חישוב.

בשירות שהמשתמשים שלו נעים בין מחשבים ניידים מדגמי הדגל לטלפונים בני עשור, תקציב הסיבוכיות הוא שמכריע. קודקים נוירוניים נכנסים לשימוש קודם היכן שבסיס המכשירים המותקן חדש, והספק שולט בשני הקצוות.

מה זה אומר לעסקים בתחום התקשורת

שיחות ועידה ארגוניות, נוכחות מרחוק, סיוע מרחוק, קווי הזנה ורשתות פרטיות הם המקומות שבהם קודק נוירוני מקצה לקצה הוא מוצר כבר היום, בתנאי ששני הקצוות שלכם. תכננו שלושה דברים מההתחלה: קודק גיבוי, בדיקות פענוח על כל NPU שאתם תומכים בו, והערכת איכות על התוכן שלכם ולא על מערך הנתונים של ספק.



פרק 05

בינה מלאכותית יוצרת

05 בינה מלאכותית יוצרת: הפריסה הגדולה ביותר, שאף אחד לא קורא לה קודק

כל מודל קול ווידאו רץ על קודק נלמד

הפריסה הגדולה ביותר של דחיסה נוירונית מקצה לקצה אינה נמצאת באף תקן קודק. היא נמצאת בתוך הבינה המלאכותית היוצרת. מודל דיבור אינו מאזין לצורות גל ואינו מפיק אותן ישירות: הוא קורא וכותב את הטוקנים הבדידים של קודק שמע נלמד. Mimi של Kyutai, שיצא ב־2024, הופך דיבור לטוקנים ב־12.5 פריימים לשנייה וב־1.1 kbps, פחות מכל קודק טלפוניה, ומשלב מידע סמנטי ואקוסטי בזרם אחד. EnCodec של Meta, שפורסם ב־2022, הוא דוגמה נוספת; המהנדסים של Meta עצמה תיארו מאוחר יותר קודקים כאלה כמצוינים בקצבי סיביות נמוכים אבל כבדים מדי לטלפונים ישנים (פרק 4).

יצירת וידאו עובדת באותו אופן. מודל דיפוזיה או טרנספורמר מייצר במרחב הלטנטי של מקודד אוטומטי נלמד לווידאו (VAE), והמפענח של ה־VAE הופך את התוצאה לפיקסלים. ה־VAE של LTX-Video דוחס וידאו ביחס 1:192: כל טוקן לטנטי מייצג בלוק של 32×32 פיקסלים לאורך 8 פריימים, עם 128 ערוצים. המפענח שלו גם מבצע את שלב הסרת הרעש האחרון, כך שהוא בחלקו מייצר פרטים ולא רק משחזר אותם.

למה אלה אינם פורמטים להפצה

הקודקים האלה רצים בכל בקשה למודל קול או וידאו, אבל אף אחד מהם אינו פורמט שמישהו אחר מפענח. לכל מודל מרחב לטנטי משלו; אין בקרת קצב, אין בדיקת התאמה ואין ערובה שמפענח אחר יפיק את אותה תמונה. הם מתוכננים לאיכות היצירה, לא לנאמנות בקצב סיביות נתון, והם מתפענחים על GPU במרכזי נתונים. ההצלחה שלהם מראה שדחיסה נלמדת עובדת בהיקף רחב כשחברה אחת מחזיקה במקודד, במפענח ובחומרה.

מה זה אומר לעסקים

שתי נקודות מעשיות. ראשית, היעילות של הטוקנייזר היא שורת עלות ישירה לכל מי שמספק שירותי יצירת וידאו או קול: בטרנספורמר שמפעיל קשב על כל הטוקנים, החישוב גדל מהר יותר ממספר הטוקנים, ולכן ייצוג לטנטי דחוס יותר מוריד את עלות היצירה יותר מאשר ביחס ישר **ניתוח** . מהנדסי קודקים שמבינים בתכנון ייצוגים לטנטיים, במודלי אנטרופיה וביעילות מפענחים הם נכס לצוותי בינה מלאכותית יוצרת. שנית, הגבול זז: גופי התקינה החלו לעבוד על קידוד למכונות ולא לבני אדם.

Video Coding for Machines של MPEG (ISO/IEC 23888-2) הגיע לשלב טיוטת תקן בינלאומי בתחילת 2026 ומתוכנן להיות סופי בסוף 2026, ו־JPEG AI תוכנן מלכתחילה לשרת משימות ראייה ממוחשבת לצד צפייה אנושית.

פרק 06

משחקים וגרפיקה

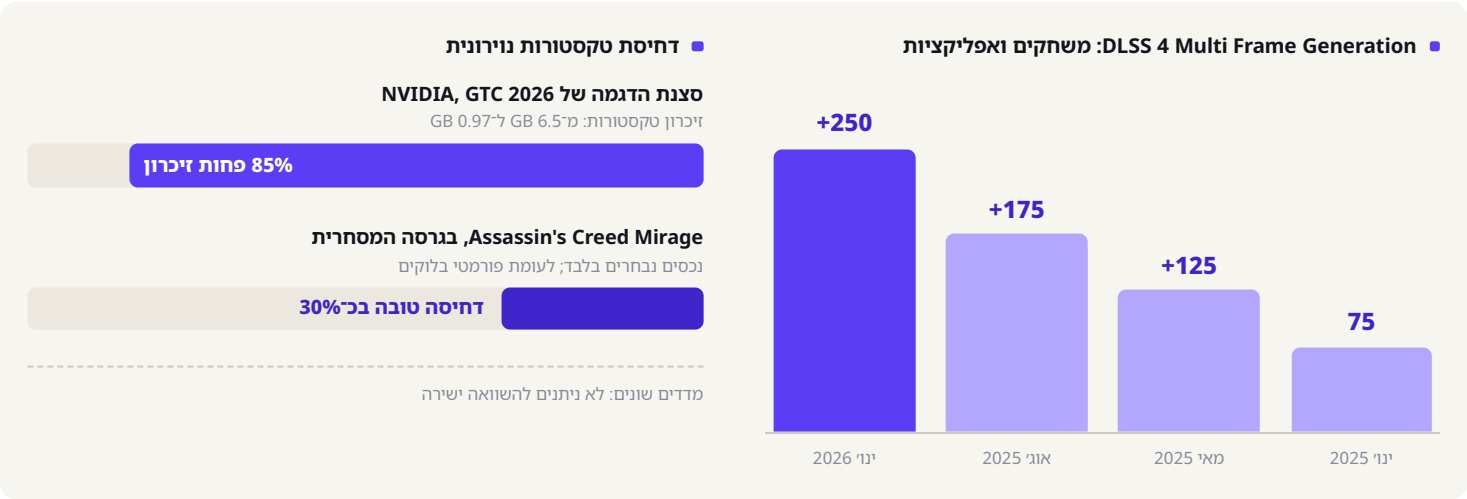
06 משחקים וגרפיקה: שחזור נוירוני כבר נפוץ, נכסים נוירוניים רק מתחילים

לרנדר פחות פיקסלים ולתת לרשת להשלים

בגרפיקה בזמן אמת, רשת כבר משחזרת את רוב מה שהשחקנים רואים. DLSS של NVIDIA מרנדר פחות פיקסלים ופחות פריימים ומשתמש ברשתות כדי לייצר את השאר. DLSS 4 עם Multi Frame Generation הושק בינואר 2025 עם תמיכה ב־75 משחקים ואפליקציות; NVIDIA ספרה יותר מ־125 עד מאי 2025, יותר מ־175 עד אוגוסט ויותר מ־250 עד ינואר 2026, אז הכריזה על DLSS 4.5 עם מודל טרנספורמר מהדור השני לרזולוציית־על. זהו מפענח לכל דבר מלבד השם: "זרם הסיביות" הוא פריים ברזולוציה נמוכה ועוד נתוני תנועה, והרשת בונה מחדש את התמונה המלאה.

איור 05 כשחברה אחת שולטת בכל השכבות, שחזור נוירוני מתפשט מהר

מימין: משחקים ואפליקציות שתומכים ב־DLSS 4 Multi Frame Generation של NVIDIA. משמאל: דחיסת טקסטורות נוירונית בסצנת הדגמה ובמשחק שיצא לשוק, שהשתמש בה רק בנכסים נבחרים בגלל עלות זמן הריצה.



מקור: NVIDIA (ינואר 2025 עד ינואר 2026; הדגמה ב־GTC 2026); Ubisoft La Forge (פברואר 2026). טענות החברות; שני נתוני הטקסטורות מודדים דברים שונים.

טקסטורות נוירוניות כבר יצאו לשוק, באופן סלקטיבי

טקסטורות נדחסות בפורמטים קבועים של בלוקים (BCn), ש־GPU מפענחים בחומרה. דחיסת טקסטורות נוירונית מחליפה סט של טקסטורות חומר בייצוג לטנטי קומפקטי וברשת קטנה שמשחזרת טקסלים כשיש בהם צורך. בפברואר 2026 תיארה מעבדת La Forge של Ubisoft איך הוציאה אותה לשוק ב־Assassin's Creed Mirage, משחק מ־2023, עם דחיסה אפקטיבית גבוהה בכ־30% מזו של הפורמטים הקונבנציונליים באותה איכות חזותית **טענת החברה**. היא לא שימשה בכל מקום: "השימוש הוגבל על ידי ביצועי זמן הריצה", ולכן Ubisoft החילה אותה על תת־קבוצה של נכסים עם מספר מופעים גבוה ולחץ זיכרון כבד.

ה־RTX Neural Texture Compression SDK של NVIDIA, בגרסת בטא ב־GitHub, מראה את התקרה. בהדגמה ב־GTC 2026, סצנת וילה שהשתמשה ב־6.5 GB של זיכרון טקסטורות עם BCn נזקקה ל־970 MB עם טקסטורות נוירוניות, 85% פחות

טענת החברה . ה־SDK מציע שני נתיבים: פענוח בזמן הטעינה ל־BCn, שחוסך מקום בדיסק ונפח הורדה אבל לא זיכרון וידאו, או פענוח בזמן הדגימה, שחוסך את הזיכרון אבל עולה זמן שיידר בכל פריים.

ה־API מגיע

ממשקי ה־API הגרפיים מוסיפים את פעולות המטריצה שפענוח נוירוני צריך בתוך שיידרים. Microsoft שחררה את Shader Model 6.9 בפברואר 2026; יכולת ה־cooperative vector שהציגה בתצוגה מקדימה ב־2025 שולבה בתכנון אחוד של אלגברה לינארית למודל שיידרים מאוחר יותר, ו־DirectX Linear Algebra נכנס לתצוגה מקדימה ציבורית באפריל 2026. כשזה יהיה בכל דרייבר, מפענח טקסטורות נוירוני יוכל לרוץ על GPU של כל יצרן מנתיב קוד אחד.

מה זה אומר לעסקים

משחקים כבר כוללים פענוח נוירוני כי התנאים של מעגל סגור מתקיימים: מנוע אחד, צינור נכסים אחד, חומרת טנזורים ב־GPU ופורמט גיבוי. העבודה הפתוחה היא בכלים, בצינורות נכסים שמקודדים חומרים נוירוניים, בסביבות ריצה חוצות־יצרנים, ומעל הכול במובייל, שבו עלות האנרגיה לכל טקסל מפוענח היא שמכריעה.



פרק 07

07

ערוצים ייעודיים

ערוצים ייעודיים: לוויינים, רחפנים ורובוטים

אלה המקרים שבהם קידוד נוירוני מקצה לקצה מתאים ביותר על הנייר. בעלים אחד שולט בשני קצות הערוץ, רוחב הפס מועט ונתון לאובדן נתונים, התוכן צר מספיק כדי לאמן עליו, ותמונה נאמנה תפיסתית, או אפילו סמנטית, מספיקה לעיתים קרובות. קודקים נוירוניים עמידים לאובדן נתונים לערוצים כאלה הם תחום מחקר פעיל, ונעשה שימוש בשחזור בצד השרת של שידורים בקצב סיביות נמוך.

הם גם שוק קטן, ומוסדר יותר ויותר, בהשוואה לווידאו הצרכני **ניתוח** . בסין, הודעה מס׳ 49 לשנת 2026 של משרד המסחר ושני גופים נוספים, שפורסמה ב־19 באוקטובר ובתוקף מ־18 בנובמבר 2026, מעדכנת את פיקוח הייצוא על 13 קטגוריות של פריטים הקשורים לרחפנים. בין סעיפי הטכנולוגיה, אחד עוסק בפלטפורמות קרקעיות בלתי מאוישות שמוגדרות בין היתר לפי ספים כמו דחיסת תמונה סמנטית של 75% ושחזור תמונה בתוך 150 מילישניות, כפי שדווח בעיתונות המקצועית. הדחיסה הסמנטית עצמה הפכה לפרמטר של פיקוח ייצוא. בארצות הברית, ועדת התקשורת הפדרלית הוסיפה ב־22 בדצמבר 2025 רחפנים מתוצרת חוץ ואת רכיביהם הקריטיים ל־Covered List שלה, כך שדגמים זרים חדשים כבר אינם יכולים לקבל אישור ציוד.

המסקנה: נישה מוצקה לחברה מתמחה עם לקוחות תעשייתיים או ממשלתיים ויכולת לנהל ציות לפיקוח ייצוא משני הצדדים, אבל לא שוק רחב, ולא המקום ללמוד אם קידוד נוירוני עובד בהיקף רחב.



סיליקון

08 המבט מהסיליקון: אנרגיה, חישוב, זיכרון ודיוק מוחלט

זה הפרק שבו הרקע שלי חשוב יותר מבכל מקום אחר. מאחוריי שנים רבות של מו״פ וניהול בתחום השבבים במודל fabless, כולל תכנון full-custom של נתיבי נתונים ב־ASIC ופרויקטים של שבבי מצלמה למולטימדיה, כמו מאיצים לקודקי וידאו, והשאלות שצוות שבבים שואל על קודק חדש אינן עוסקות ב־BD-rate. הן: כמה ג׳אולים לפיקסל, כמה בתים של תעבורת DRAM לפיקסל, כמה שערים, והאם כל מפענח בעולם יכול לחשב בדיוק את אותן סיביות? בכל הארבע, מפענח נוירוני ב־2026 נמצא היכן ש־HEVC בחומרה היה לפני השבבים הראשונים שלו: מבטיח בתוכנה, יקר בסיליקון.

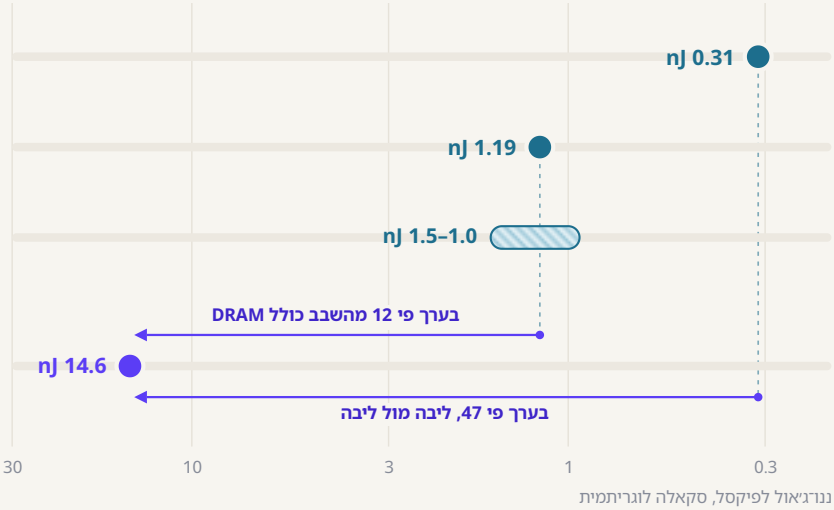
אנרגיה לפיקסל: הבדל של סדר גודל

מעט מפענחים נוירוניים תוכננו כשבבים, ולכן ההשוואה נשענת על קומץ תכנונים שפורסמו. נקודת הייחוס בצד הקלאסי היא שבב מפענח ה־HEVC של MIT (Tikekar et al., ISSCC 2013 ו־IEEE JSSC 2014): 40 nm, 1.77 mm², מפענח 4K ב־30 פריימים לשנייה ב־200 MHz. הליבה שלו צורכת 78 mW, כלומר 0.31 nJ לפיקסל; כשסופרים גם את ההספק המדומה של ה־DRAM, המערכת צריכה 1.19 nJ לפיקסל. עבור VVC, נתוני הייצור קניניים; ניתוח ב־Electronic Design מ־2026, שממפה את שבב MIT לתהליך של 28 nm ומחיל עליו את הסיבוכיות הגבוהה יותר של VVC, צופה 1–1.5 nJ לפיקסל למפענח 4K60 של כמיליון שערים.

בצד הנוירוני, תכנון ה־ASIC היחיד שפורסם למפענח וידאו נוירוני מקצה לקצה, לפי הספירה של Electronic Design, הוא NVCA (Zhang et al., DATE 2024), שסונתז בתהליך 28 nm של TSMC עם 5.01 מיליון שערים, בלי לספור את המשקלים שלו, שנמצאים מחוץ לשבב. הוא מפענח 1080p ב־25 פריימים לשנייה ב־760 mW, כלומר 14.6 nJ לפיקסל.

איור 06 אנרגיה לפיקסל מפוענח: מפענחים ייעודיים מובילים בסדר גודל

תכנון השבב היחיד שפורסם למפענח וידאו נוירוני צורך לפיקסל בערך פי 12 אנרגיה משבב מפענח HEVC שנמדד, כשסופרים גם את ה־DRAM שלו, ובערך פי 47 מהליבה שלו לבדה. תעבורת המשקלים וה־DRAM של התכנון הנוירוני עצמו אינה נספרת.



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.

מקור: Tikekar et al., IEEE JSSC 2014 (MIT); NVCA, Zhang et al., DATE 2024 (arXiv 2312.10716); הטווח של VVC הוא תחזית של Electronic Design (2026). שבב ה־HEVC מיוצר בתהליך ישן יותר (40 nm לעומת 28 nm), מה שמחמיא לתכנון הנוירוני.

זה בערך פי 12 משבב ה־HEVC כשסופרים את ה־DRAM שלו, ובערך פי 47 מהליבה שלו לבדה. ההשוואה מחמיאה ל־NVCA בשלושה אופנים: שבב ה־HEVC מיוצר בתהליך ישן יותר, הנתון של NVCA אינו כולל את תעבורת המשקלים וה־DRAM שלו, ו־NVCA הוא תוצאת סינתזה ולא סיליקון שנמדד. ב־1080p30 ההבדל הוא כ־74 mW לעומת כ־900 mW: מפענח שאף אחד לא מרגיש, לעומת מפענח שמורגש בסוללה ובתקציב החום של הטלפון.

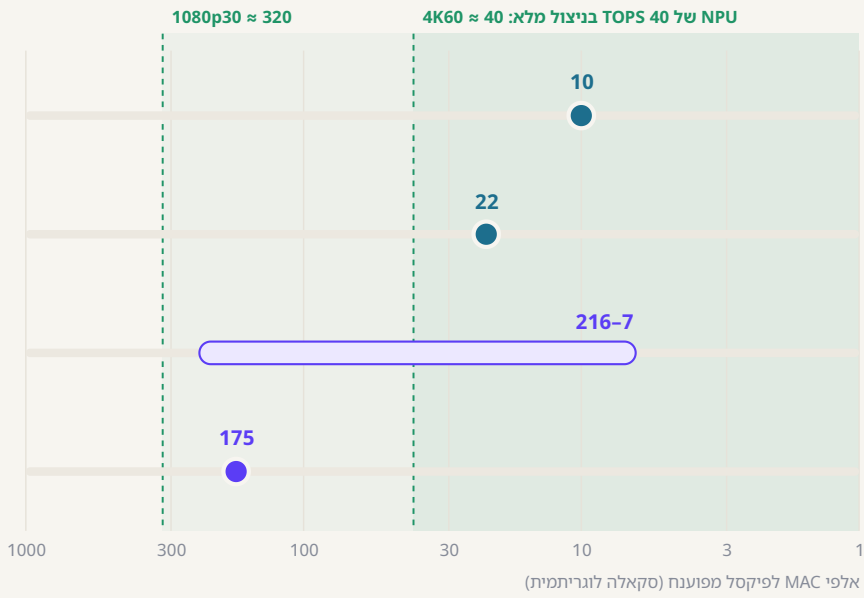
חישוב לפיקסל: מה מפענח נוירוני דורש מ־NPU

קודק הווידאו הנוירוני המהיר ביותר שפורסם ושמצאתי הוא DCVC-RT של Microsoft Research (Jia et al., CVPR 2025). הוא מקודד ומפענח 1080p ב־125 וב־113 פריימים לשנייה, בהתאמה, על GPU מסוג NVIDIA A100, תוך חיסכון של 21% בקצב הסיביות לעומת מקודד הייחוס של VVC בתנאי הבדיקה של המאמר, והוא משתמש באריתמטיקה של מספרים שלמים כדי שמכשירים שונים יפענחו באופן עקבי. החומר המשלים שלו מפרט את הסיבוכיות של כל מודול. כשאני מחבר את ארבעת המודולים שמפענח חייב להריץ (מפענח, מחלץ מאפיינים, מודל אנטרופיה ושחזור), אני מקבל כ־175 אלף פעולות הכפלה־וצבירה לכל פיקסל מפוענח אומדן.

ב־1080p30 מדובר בכ־11 טריליון פעולות הכפלה־וצבירה בשנייה, כלומר כ־22 TOPS בחישוב של שתי פעולות ל־MAC, לפני כל אובדן ניצולת; 4K60 דורש פי שמונה אומדן. לשם השוואה, Microsoft מגדירה את מחשבי ה־AI שלה לפי NPU של 40 TOPS לפחות, ו־Apple מדרגת את ה־Neural Engine של M4 ב־38 TOPS. אלה נתוני שיא, שמתחלקים עם כל שאר הדברים שהמכשיר עושה.

איור 07 **כלי נוירוני עולה עשרות אלפי MAC לפיקסל; מפענח נוירוני שלם עולה מאות אלפים**

פעולות הכפלה־וצבירה של המפענח לפיקסל, בסקאלה לוגריתמית, לעומת מה ש־NPU של 40 TOPS בניצול מלא יכול להרשות לעצמו ב־4K60 וב־1080p30. השיפורים נמדדים מול נקודות ייחוס שונות ואינם ניתנים להשוואה ישירה.



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
27.9%–12.5 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

מקור: JVET דרך יו״ר SC 29 (אפריל 2026); ועדת JPEG (2024), החומר המשלים ל־DCVC-RT, סכום מודולי המפענח שחישב המחבר; CVPR 2025, ההגדרה של Microsoft למחשבי Copilot+. החישוב של המחבר: שתי פעולות ל־MAC, ניצול מלא.

הסולם באיור 7 הוא הדרך השימושית לקרוא את התחום. כלים נוירוניים שעולים עשרות אלפי MAC לפיקסל, כמו מסנני הלולאה של JVET (חיסכון של כ־6% ב־10 אלף, 8% ב־22 אלף), סבירים בתוך מפענח תקני עתידי. JPEG AI פרוש על פני כל הטווח, מ־7 עד 216 אלף MAC לפיקסל, עבור 12.5–27.9% לעומת VVC תוך־תמונתי. מפענח וידאו נוירוני שלם, בכמאתיים אלף MAC לפיקסל, מקומו על NPU, ברזולוציה צנועה, במעגל סגור. בדיוק שם נמצא MLVC: 540p30, עם ניצול של לכל היותר מחצית מה־NPU.

תעבורת זיכרון, לא MAC, היא לעיתים קרובות המגבלה האמיתית

מחברי DCVC-RT מצאו שמה שהגביל את המהירות לא היה החישוב אלא "העלות התפעולית": קלט ופלט של זיכרון ומספר קריאות הפונקציה. הם הסירו מודולי תנועה מפורשים ועיבוד רב־קנה־מידה בעיקר מהסיבה הזו. המתכננים של NVCA בנו זרימת נתונים מיוחדת כדי לצמצם את תעבורת הזיכרון שמחוץ לשבב, שנובעת ממאפייני התנועה והשארית. מתכנן שבבים יזהה את הדפוס. בשבב ה־HEVC של MIT, ה־DRAM אחראי לכשלושה רבעים מאנרגיית המערכת (0.88 מתוך 1.19 nJ לפיקסל), אפילו אחרי מטמון שהפחית את רוחב הפס של ה־DRAM ב־67% **ניתוח**. מפענח נוירוני ששומר מפות מאפיינים גדולות ורב־ערוציות בין פריימים מחמיר את בעיית הזיכרון, לא משפר אותה.

דיוק מוחלט: כל מפענח חייב לחשב את אותן סיביות

קידוד אנטרופיה עובד רק אם מודל ההסתברות של המפענח תואם בדיוק את זה של המקודד. בקודק קלאסי המודל הזה הוא טבלה או קבוצה קטנה של כללים במספרים שלמים. בקודק נוירוני הוא רשת, והבדל בנקודה צפה בין שני מכשירים, בגלל קרנל אחר, סדר צבירה אחר או NPU אחר, יכול לגרום למפענח לקרוא סמלים שגויים ולאבד את הזרם מאותה נקודה והלאה.

כל מי שהוציא לשוק או תקנן קודק נוירוני נאלץ לפתור את זה בהנדסה. DCVC-RT ממיר את המודל שלו למספרים שלמים לקידוד עקבי בין מכשירים. MLVC משתמש בתכנון hyperprior עם קנה מידה משותף כדי לשרוד אי־התאמה בין פלטפורמות, וזו כנראה חלק מהסיבה שהוא מוותר על דחיסה לעומת DCVC-RT **ניתוח**. ניסויי רשתות הנוירונים של JVET מריצים כל רשת דרך SADL, ספרייה קטנה להסקה במספרים שלמים של 16 סיביות. JPEG AI מגדיר נקודות התאמה בתחום הלטנטי, בתחום

השארית ובתחום הפיקסלים. לצוות שבבים זו אותה משמעת כמו בהתמרות המספרים השלמים של HEVC: התקן חייב להגדיר את האריתמטיקה במדויק, מה שמגביל את תכנון הרשת ופוסל קיצורי דרך ייחודיים לספק.

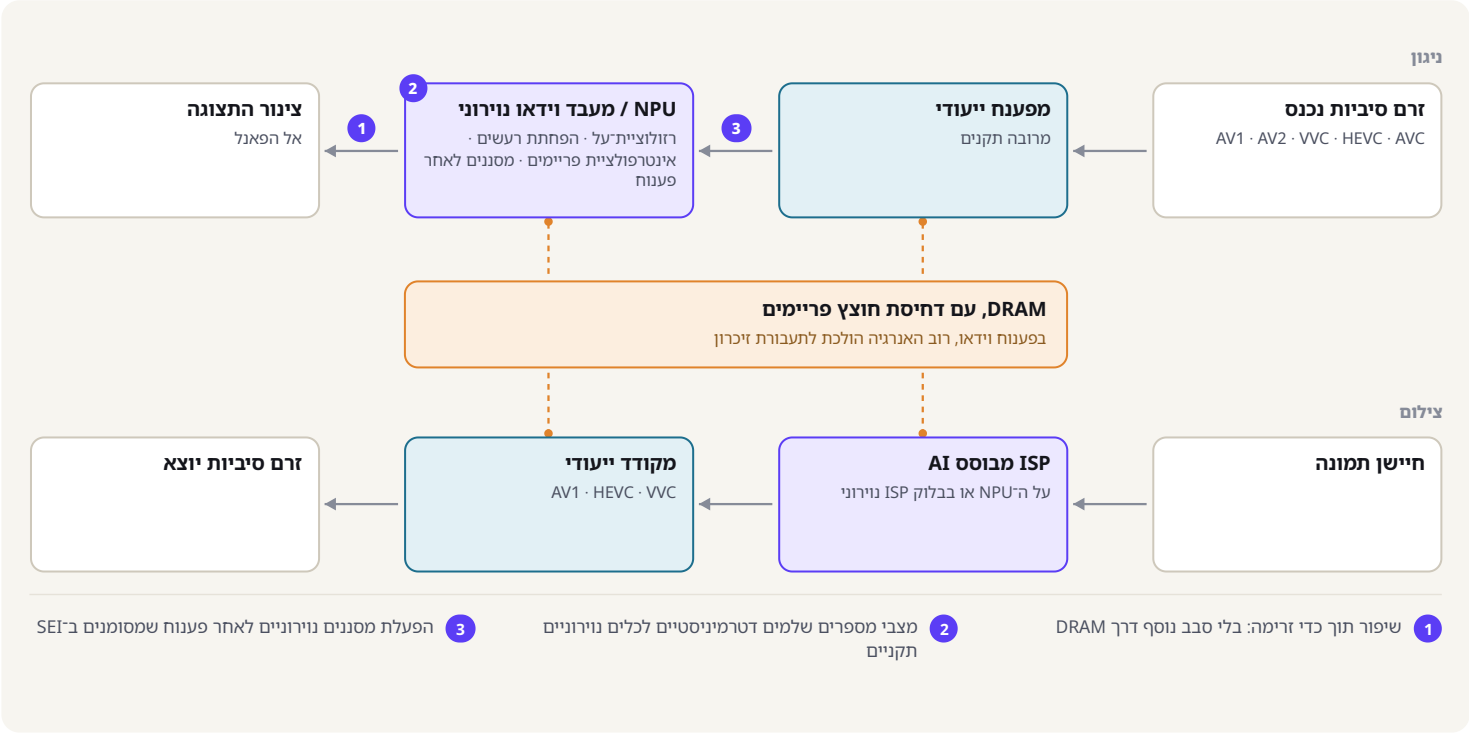
השטח אינו המכשול; האימות ומשך החיים הם המכשול

5 מיליון השערים של NVCA, לעומת כמיליון למפענח VVC לפי תחזית, נראים הרבה, אבל ב־SoC מודרני של טלפון או טלוויזיה שכבר נושא NPU של כמה TOPS, כמה מיליוני שערים אינם מה שמכריע את התכנון. מה שעולה כסף אמיתי הוא אימות מול חבילת בדיקות התאמה, שמירה על נכונות של מפענח שניתן לתכנות לאורך עדכוני תוכנה במשך כל חיי המכשיר, וההספק ב־4K60 HDR. העלויות האלה הגיוניות רק אחרי שתקן מקבע מה צריך לבנות.

הארכיטקטורה שמנצחת עכשיו: קודק ייעודי ולצדו NPU

איור 08 **החלוקה שבשטח: חומרה ייעודית לזרם הסיביות, מנוע נוירוני לפיקסלים**

SoC מדיה של 2026. בלוק הקודק מפענח כל תקן בשבריר ננו־ג׳אול לפיקסל; ה־NPU עושה את מה שרשתות טובות בו ושהתקנים אינם מגבילים. תעבורת ה־DRAM היא העלות שקובעת.



מקור: תפיסת הארכיטקטורה של המחבר, על סמך פרסומי מוצר של Tencent (Canghai V2), Chips&Media, Allegro DVT ו־Samsung (2026).

הסיליקון שנמכר ב־2026 משלב קודק ייעודי מרובה תקנים עם מנוע נוירוני לצדו. בלוק הקודק מטפל ב־AV1, AV2, VVC, HEVC ו־AVC בשבריר ננו־ג׳אול לפיקסל. המנוע הנוירוני מטפל במה שרשתות טובות בו ושהתקנים אינם מגבילים: הפחתת רעשים וה־ISP בצד הצילום, ורזולוציית־על, אינטרפולציית פריימים ושיפור בצד התצוגה. השניים חולקים את דחיסת חוצץ הפריימים ואת תזמון ה־DRAM, כי תעבורת הזיכרון היא העלות שקובעת.

השוק מראה את החלוקה. ה־ASIC לקודק וידאו Canghai (沧海) של Tencent, שמפותח בבית מאז 2019, נמצא בפריסה של יותר מ־100,000 שבבים מהדור הראשון ב־Tencent Cloud ובשירותי השידור החי, הווידאו לפי דרישה ומשחקי הענן של החברה עצמה; הדור השני, שעליו הוכרז במאי 2026 כשבב שכבר הודלק ונכנס לייצור המוני לקראת הפעלה במחצית השנייה של השנה, תומך ב־H.265 וב־H.266 בארכיטקטורה אחת, עם דחיסה טובה ביותר מ־10% וביותר מ־30% בהתאמה לעומת קודמו, בערך פי שניים תפוקה, ושיפור תמונה על השבב כמו רזולוציית־על. Chips&Media מוכרת IP של קודקים ו־NPU מוכוון וידאו, WAVE-N, זה לצד זה. Allegro DVT מוכרת מפענחי AV2 והוסיפה את ה־IP הראשון שלה לעיבוד וידאו נוירוני. מעבדי הטלוויזיה של Samsung משלבים מפענחים תקניים עם מאות רשתות להגדלת רזולוציה.

מה הייתי בונה, לפי תפקיד

צוותי SoC ושבבי טלוויזיה. שמרו על מפענחים ייעודיים לכל תקן שאתם חייבים לנגן, כולל AV2 כבר עכשיו ו־VVC היכן שהשידור מחייב זאת. התאימו את גודל ה־NPU לעיבוד לאחר פענוח ברזולוציה ובקצב הפריימים של התצוגה, ותכננו את הנתיב מהמפענח ל־NPU ולתצוגה כך שהשיפור לא יוסיף סבב דרך ה־DRAM. תמכו ב־SEI של מסננים נוירוניים לאחר פענוח. ספקו מצבי מספרים שלמים דטרמיניסטיים ב־NPU. התייחסו ל־H.267 כתחילת תכנון ב־2029 או אחר כך, ושמרו על מידה של יכולת תכנות עבור הכלים הנוירוניים שלו.

ספקי IP. מכרו עכשיו מפענחי AV2 ו־VVC, ובנוסף את הרכיבים שעיבוד נוירוני צריך סביבם: מאיצים למסננים לאחר פענוח, דחיסת חוצץ פריימים וכלי בדיקת התאמה.

פלטפורמות וידאו. עיבוד מקדים נוירוני במקודד מחזיר את ההשקעה ראשון, כי הוא רץ פעם אחת לכל כותר ואינו דורש דבר מהמכשירים. השתמשו ברזולוציית־על בצד הלקוח היכן שהמכשירים מאפשרים זאת. שמרו את הקודקים הנוירוניים מקצה לקצה למעגלים סגורים שבשליטתכם.

תקנים ופטנטים

09 תקנים ופטנטים: היכן כלים נוירוניים נעשים ניתנים לרישוי

צנרת התקנים

תקן	גוף	תוכן נוירוני	מצב, נובמבר 2026
H.267 (שם עבודה; מעבר ל־VVC)	JVET (ITU-T ו־ISO/IEC)	פתוח. כלים קלאסיים (ECM) וכלים נוירוניים (מסנני לולאה, חיזוי תוך־תמונתי, רזולוציית־על) מועמדים שניהם	קול קורא להצעות ביולי 2026; הערכת המענים בינואר 2027; מודל בדיקה ראשון במחצית השנייה של 2027; גרסה 1 בסוף 2029
VVC ו־HEVC, עם VSEI	JVET	אפשר לתאר מסנן נוירוני לאחר פענוח במטא־נתוני SEI; הפעלתו אופציונלית למפענח	בתוקף; ההודעות הורחבו במהדורות אוקטובר 2025
AV2	AOMedia	כלים קלאסיים, ללא תמלוגים	גרסה 1.0.0 ב־28 במאי 2026; IP של מפענחים נמכר ברישיון מיוני 2026; סטרימינג של YouTube באתר מתוכנן לתחילת 2027
JPEG AI	JPEG (ISO/IEC ו־ITU-T)	קודק תמונה נלמד מקצה לקצה, לראייה אנושית ולראיית מכונה	חלק 1 פורסם ב־2025; גרסה שנייה בדיון; לא נמצאה תמיכה כברירת מחדל במוצרים
IEEE 1857.11	IEEE	קידוד תמונה מבוסס רשתות נוירונים שתוכנן ל־NPU	פורסם בדצמבר 2024
Video Coding for Machines	MPEG	קידוד לניתוח ממוחשב ולא לצפייה	טיוטת תקן בינלאומי בתחילת 2026; הגרסה הסופית מתוכננת לסוף 2026
הרחבות ה־ML של Opus	IETF	יתירות עמוקה, שיפור בצד המפענח והרחבת רוחב פס, כולם אופציונליים	ב־libopus 1.5 ו־1.6; היתירות העמוקה בתהליך תקינה

שני דברים בולטים. ראשית, תוכן נוירוני נכנס לתקנים בשוליים לפני שהוא נכנס לליבה: כמסננים אופציונליים לאחר פענוח, כשיפור אופציונלי וכקידוד תמונה, שכולם משאירים את נתיב הפענוח הראשי ללא שינוי. שנית, התקן היחיד שיכול להכניס כלים

נוירוניים לנתיב הפענוח של כל מסך, H.267, רחוק שלוש שנים מגרסה ראשונה.

מי ממוצב בפטנטים

כלים בתקן הופכים לפטנטים שאפשר למכור עליהם רישיונות, ולכן תהליך H.267 הוא גם מרוץ פטנטים, והכלים הנוירוניים הם חלק ממנו. מיפוי הפטנטים של GreyB מ־2025, שכולל 3,916 פטנטים מעבר ל־VVC, מציב את ByteDance בראש, עם 645 משפחות פטנטים ו־251 תרומות ל־JVET, ומציין שותפויות בין אקדמיה לתעשייה, כמו Xiaomi ואוניברסיטת Xidian, עם יותר מ־20 משפחות. MLex דיווחה במרץ 2026 ש־Tencent, Alibaba ו־ByteDance מתבלטות כמשתתפות פעילות במאגרי פטנטים של קודקי וידאו. במאי 2026 קנתה Sauvegarder Investment Management מ־Alibaba חבילה של 126 נכסי פטנט של קודקי וידאו, שמכסים את AV1, AV2 ו־Video Coding for Machines. הרכישה של Deep Render על ידי InterDigital באוקטובר 2025 הוסיפה לתיק שלה את הפטנטים של Deep Render בקידוד וידאו מבוסס AI.

את הכסף לכל מכשיר קל לראות בטבלאות התעריפים של המאגרים. התעריפים של Access Advance עבור VVC (נכון ל־1 ביולי 2026, לבעלי רישיון שעומדים בתנאים ומשתמשים בסימן המסחרי שלה) הם 0.50 דולר לטלפון נייד, טאבלט או מחשב נייד בשווקים של אזור 1 כמו ארה״ב, האיחוד האירופי ויפן, ו־0.25 דולר באזור 2; לטלוויזיות 4K, 1.00 עד 1.50 דולר לפי המחיר באזור 1; עם תקרה שנתית לארגון של 75 מיליון דולר. AV1 ו־AV2 פטורים מתמלוגים מצד חברות AOMedia. מה שלא יכיל H.267, הכלים שלו יתומחרו באותו אופן, ולכן חברות עם פטנטים בקידוד נוירוני תורמות ל־JVET כבר עכשיו ניתוח.

פרק 10

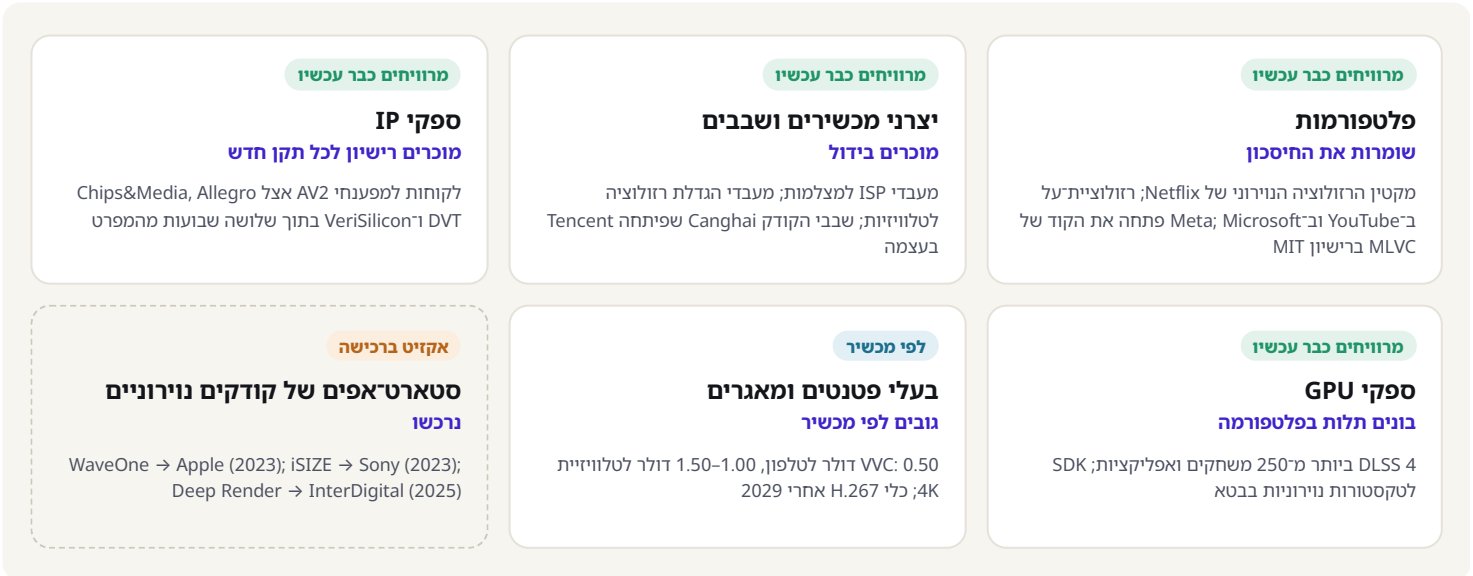
10 מי מרוויח, והיכן חברה יכולה לפעול

מי מרוויח

הכסף ב־2026

איור 09 **מי מרוויח מקידוד נוירוני ב־2026**

הכסף הולך בעקבות התקנים והפלטפורמות. לא מצאתי אף חברה עצמאית שמתפרנסת מרישוי של קודק וידאו נוירוני מקצה לקצה.



מקור: פרסומי חברות ומאגרי פטנטים שמצוטטים בדוח; המצב נכון ל־12 בנובמבר 2026. תעריפי המאגר: VVC, Access Advance, אזור 1, בעמידה בתנאים ועם סימן מסחרי, נכון ל־1 ביולי 2026.

הפלטפורמות שומרות את החיסכון. Netflix, YouTube, Meta, Microsoft ו־Tencent בונות עיבוד נוירוני בבית ושומרות לעצמן את התועלת: פחות רוחב פס, פחות אחסון ואיכות טובה יותר לכל סיבית. כשזה עוזר לאימוץ, הן מוסרות את הקוד: Microsoft שחררה את MLVC ברישיון MIT.

יצרני המכשירים והשבבים מוכרים בידול. מעבדי ISP במצלמות ומעבדי תמונה בטלוויזיות הם המקום שבו רשתות נוירונים נוגעות בכל משתמש, והם נמכרים כתכונות: ה־Cognitive ISP של Qualcomm, מעבדי הגדלת הרזולוציה של Samsung.

ספקי IP מוכרים מפענחים לכל תקן חדש, ובנוסף NPU. בתוך שלושה שבועות מפרסום המפרט של AV2, הכריזו Chips&Media, Allegro DVT ו־VeriSilicon כולן על לקוחות למפענחי AV2. אותם ספקים מוכרים כיום בלוקים לעיבוד נוירוני לצד הקודקים.

ספקי GPU בונים פלטפורמות. DLSS ודחיסת טקסטורות נוירונית הופכים את חומרת הטנזורים של NVIDIA לחלק מהאופן שבו משחקים נבנים.

בעלי פטנטים גובים לפי מכשיר, דרך מאגרים ל־VVC ול־HEVC, ואחרי 2029 גם ל־H.267.

סטארט־אפים של קודקים נוירוניים נרכשו. WaveOne עברה ל־Apple ב־2023, iSIZE ל־Sony Interactive Entertainment ב־2023, ו־Deep Render ל־InterDigital ב־2025, כולן בתנאים שלא פורסמו. בכל אחד מהמקרים, פלטפורמה או מעניק רישיונות קנו את הצוות ואת הפטנטים. לא מצאתי ב־2026 אף חברה עצמאית שמתפרנסת מרישוי של קודק וידאו נוירוני מקצה לקצה.

היכן חברה יכולה לפעול

הזדמנות	מי קונה	למה עכשיו	הסיכון העיקרי
עיבוד מקדים נוירוני במקודד (הקטנת רזולוציה, הפחתת רעשים, כיוונון לכל כותר)	פלטפורמות סטרימינג ווידאו של משתמשים, גופי שידור	אין צורך בשינוי במכשירים; הוכח ב־Netflix	פלטפורמות גדולות בונות זאת בעצמן
IP וכיוונון לשיפור בצד התצוגה (רזולוציית־על, אינטרפולציית פריימים, איכות תמונה)	יצרני שבבים לטלוויזיות, לממירים ולטלפונים	ה־NPU כבר נמצא במכשירים; בידול נראה לעין	הפיכה למוצר מדף; פיתוח פנימי אצל יצרני השבבים
קודקים נוירוניים במעגל סגור (שיחות ועידה, הפעלה מרחוק, ערוצים פרטיים)	ספקי תקשורת ארגונית, תעשייה ונוכחות מרחוק	MLVC מוכיח היתכנות, והקוד שלו פתוח	חייבים לשלוט בשני הקצוות; פיצול בין NPU שונים
סיליקון או IP של קודק ו־NPU (AV2, VVC ועיבוד לאחר פענוח)	יצרני SoC	מחזור AV2 מתחיל	שחקנים ותיקים וחזקים
כלים להערכה, לבדיקות התאמה ולהסקה 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

לרוב החברות, שתי השורות הראשונות הן המקום שבו יש הכנסות זמינות ב־2027. השלישית היא המעניינת ביותר לחברת מוצר ממוקדת. האחרונה דורשת סבלנות ואסטרטגיית רישוי.

פרק 11

טענות וראיות

11 טענות שתשמעו, ומה הראיות אומרות

טענה	מה הראיות אומרות
"קודקים נוירוניים כבר מנצחים את VVC."	במעבדה, כן. DCVC-RT חוסך 21% לעומת מקודד הייחוס של VVC ב־1080p בתנאי המאמר שלו; JPEG AI חוסך 12.5–27.9% לעומת VVC תוך־תמונתי. המפענח עולה כ־175 אלף MAC לפיקסל, או את האנרגיה של NPU, לעומת שבריר ננו־ג׳אול בחומרה ייעודית.
"קודקי AI יחליפו את H.26x ואת AV1 בתוך כמה שנים."	לא בניגון. התקן הראשון שיכול לשאת כלים נוירוניים מכוון לסוף 2029, השבבים יבואו בסביבות 2031 והשימוש הרחב באמצע שנות ה־2030 אומדן. הקודק שמגיע עכשיו, AV2, הוא קלאסי.
"MLVC חוסך 88% מרוחב הפס."	לעומת H.264 ב־360p, על מערך הבדיקה של Microsoft עצמה. לעומת HEVC בחומרה מדובר ב־75.5% ב־360p וב־65.4% ב־540p, בשיחות שבהן Microsoft שולטת בשני הקצוות.
"טקסטורות נוירוניות חוסכות 85–96% מזיכרון הווידאו."	בסצנות הדגמה. המשחק שיצא לשוק השיג דחיסה אפקטיבית טובה בכ־30%, והשתמש בטקסטורות נוירוניות רק בנכסים נבחרים בגלל עלות זמן הריצה.
"JPEG AI הוא תקן ISO, ולכן מצלמות ודפדפנים יאמצו אותו."	לא מצאתי תמיכה כברירת מחדל במוצרים. הפורמט שהאינטרנט אימץ ב־2026, JPEG XL ב־Chrome 155, הוא קלאסי.
"פענוח נוירוני יקר מדי לטלפונים."	נכון בחלקו. MLVC מפענח 540p30 בתוך מחצית NPU; 4K60 עם מודל מסוג DCVC-RT ידרוש כ־175 TOPS של חישוב רציף אומדן.
"ערוצי לוויין ורחפנים הם השוק הגדול לווידאו נוירוני."	מבחינה טכנית הם ההתאמה הטובה ביותר, אבל השוק קטן וכפוף יותר ויותר לפיקוח ייצוא משני צדי האוקיינוס השקט.
"פלטפורמות גדולות ירכשו רישיון לקודק נוירוני מסטארט־אפ."	הן בנו בעצמן (Netflix, Microsoft, Tencent) או קנו את הסטארט־אפ (Apple, Sony, InterDigital).

מבחן פשוט מנפה את רוב הפיצ׳ים: בקשו את נקודת הייחוס, המדד, הרזולוציה ועלות הפענוח. טענה שאינה יכולה לציין את כל הארבעה היא תוצאת מחקר, לא מוצר.

פרק 12

12

ציוני דרך

מה כדאי לעקוב אחריו, 2027 עד 2030

מתי	ציון דרך	למה זה חשוב
ינואר 2027	JVET מעריך בבריסביין את המענים ל־H.267	כמה מהמענים המובילים נשענים על כלים נוירוניים, ובאיזו עלות לפיקסל
תחילת 2027	הסטרימינג הראשון של AV2 באתר של YouTube; מפענחי AV2 ב־SoC שהוכרזו	המחזור הקלאסי הבא קובע את תקציבי המכשירים עד 2030 בערך
המחצית השנייה של 2027	מודל הבדיקה הראשון של H.267	האם הכלים הנוירוניים נמצאים בתכנון הליבה או רק בשלב הניסויים
2027	MLVC מעבר לשיחות עמית־לעמית ב־Teams; ספקים אחרים שמאמצים את הקוד הפתוח	המבחן הראשון של קודק נוירוני פתוח מחוץ לבעליו
2028–2027	מוצר ראשון שכותב או קורא JPEG AI כברירת מחדל	הפורמט הנוירוני הראשון שמשמש מעבר לגבולות של חברה אחת
2028–2027	DirectX Linear Algebra בדרייברים מסחריים; טקסטורות נוירוניות ביותר משחקים ובמובייל	פענוח נוירוני חוצה־יצרנים בתוך הגרפיקה
באופן שוטף	תקציבי NPU בטלפונים ובמחשבים	בכ־100 TOPS רציפים, מפענח 1080p30 מסוג DCVC-RT יתפוס כחמישית מה־NPU במקום מחצית ניתוח
סוף 2029	גרסה 1 של H.267	תחילת מחזור הסיליקון לכלים נוירוניים בניגון של השוק הרחב



13

שאלות פתוחות

שאלות פתוחות

לכמה שאלות בדוח הזה עדיין אין תשובה פומבית, ואלה השאלות שמעניינות אותי ביותר מבחינה טכנית.

דיוק מוחלט בקנה מידה רחב. עד כמה מפענח נוירוני יכול להתקרב לאריתמטיקה זהה ברמת הסיבית, במספרים שלמים בלבד, על NPU של יצרנים שונים, לפני שהוא מוותר על רוב יתרון הדחיסה שלו? הוויתור של MLVC על כשש נקודות הוא נקודת נתונים אחת; תקן צריך תשובה כללית.

זיכרון לפני MAC. כמה עולה עיבוד נוירוני לאחר הפענוח בתעבורת DRAM לפיקסל, כשהוא רץ תוך כדי זרימה בין מפענח חומרה לתצוגה, ועד כמה אפשר להקטין את סט העבודה שלו?

המחיר של כלי נוירוני ב־H.267. בכמה MAC וננו־ג׳אולים לפיקסל מסנן לולאה נוירוני מפסיק להצדיק את הסיביות שהוא חוסך במפענח 4K60?

מעגלים סגורים מעבר לשיחות. אילו מערכות סגורות אחרות, מאחסון תמונות בענן וסטרימינג של משחקים ועד רובוטים וקווי הזנה פרטיים, יכולות לאמץ קודקים נוירוניים מקצה לקצה לפני ש־H.267 יגיע?

מפענחים שמייצרים. כשמפענח מייצר פרטים במקום לשחזר אותם, כפי ש־VAE לווידאו כבר עושים, איך צריך למדוד ולהגדיר נאמנות?

אלה תחומים שנוכל לחקור יחד בשנים הבאות, דרך חילופי ידע טכני ושיתופי פעולה טכניים: i@tanshuai.com.

המגבלות של הדוח

זהו מחקר שולחני המבוסס על מסמכים פומביים עד 12 בנובמבר 2026. לא ערכתי ראיונות ראשוניים ולא מדידות בלתי תלויות למהדורה הזו. כמה מספרי מפתח הם טענות של ספקים או של מחברים על מערכי הבדיקה שלהם עצמם, כמו החיסכון בקצב הסיביות של MLVC ויחסי הדחיסה של טקסטורות נוירוניות; הם מסומנים ככאלה. אומדני החישוב שלי ל־DCVC-RT מסכמים את הנתונים לכל מודול שמחבריו מפרסמים ומתעלמים מאובדני ניצולת, כך שהעומס האמיתי על ה־NPU יהיה גבוה יותר. השוואת האנרגיה נשענת על כמה תכנונים שפורסמו בתהליכי ייצור שונים; נתוני הייצור של מפענחי VVC ו־AV2 קנייניים. לוחות הזמנים של התקנים הם היעדים של הוועדות עצמן. מועדי H.267 שאחרי 2029 הם אומדנים שלי, על סמך ההיסטוריה של AV1 ו־VVC.

על המחבר

תאן שואי הוא טכנולוג ומייסד עם יותר מעשרים שנות עבודה בתחומי השבבים, החומרה ותוכנת המערכת, ו־20 פטנטים שהוענקו לו. מאחוריו שנים רבות של מו״פ וניהול בתחום השבבים במודל fabless, כולל תכנון full-custom של נתיבי נתונים ב־ASIC ופרויקטים של שבבי מצלמה למולטימדיה, כמו מאיצים לקודקי וידאו, וזו העדשה של הדוח הזה: אנרגיה לפיקסל, תעבורת זיכרון, ומה תקן מאפשר למפענח לקבע בסיליקון. עבודתו הובילה אותו דרך עמק הסיליקון, ניו יורק, שיקגו, טייוואן, אזור המפרץ הגדול (שנזן והונג קונג), וייטנאם ומלזיה. כיום הוא עובד כחוקר ומפתח בינה מלאכותית.

לדיון טכני: i@tanshuai.com · tanshuai.com

הערות והסתייגות

הדוח הזה הוא מחקר שולחני עצמאי המבוסס על מסמכים פומביים שהיו זמינים עד 12 בנובמבר 2026. הוא מוגש למטרות מידע בלבד ואינו ייעוץ השקעות, ייעוץ משפטי או ייעוץ הנדסי. מוצרים, לוחות זמנים של תקנים, תעריפי מאגרי פטנטים וכללי ייצוא משתנים; אמתו כל דבר שאתם מסתמכים עליו מול המקור הראשוני. נתונים המסומנים טענת החברה , אומדן או ניתוח יש לקרוא בהתאם. שמות החברות והמוצרים שייכים לבעליהם ומשמשים לזיהוי בלבד; אף חברה לא סקרה את הדוח הזה ולא מימנה אותו.

© 2026 תאן שואי. מותר לשתף את הדוח הזה ללא שינוי ועם ייחוס.

מקורות

כולם נבדקו בין 3 ל־12 בנובמבר 2026. מקובצים לפי פרק.

תעבורה, מכשירים והיקף

- Ericsson, *Mobility Report*, יוני 2026 (וידאו הוא כ־75% מתעבורת הנתונים הסלולרית בסוף 2025): https://www.ericsson.com/en/reports-and-papers/mobility-report ; תקציר: https://telecomlead.com/5g/ericsson-mobility-report-2026-global-mobile-data-traffic-to-reach-515-eb-by-2031-as-ind eport/reports/june-2026 ia-leads-smartphone-data-consumption-126290
- IDC, *משלוחי סמארטפונים עולמיים, Q4 ושנת 2025 המלאה* (1.26 מיליארד), 13 בינואר 2026: https://www.idc.com/resource-center/press-releases/q425mobile /phonetop5
- TrendForce, תחזית משלוחי הטלוויזיות ל־2025 (195.71 מיליון): https://iconnect007.com/article/146406/early-shipments-by-tv-brands-push-global-1h25-vol ; Omdia, תחזית הטלוויזיות ל־2025 (208.7 מיליון): https://iconnect007.com/article/1 ume-up-2-but-raise-concerns-for-secondhalf-peak-season/146403/ein 45954/omdia-forecasts-flat-tv-shipments-in-2025-but-mini-led-poised-for-explosive-growth-at-829/145951/ein

ניגון וסטרימינג

- Netflix Technology Blog, "AV1 — Now Powering 30% of Netflix Streaming", דצמבר 2025: https://netflixtechblog.com/av1-now-powering-30-of-netflix-st reaming-02f592242d80
- Netflix Technology Blog, "For your eyes only: improving Netflix video quality with neural networks", נובמבר 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 במאי 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 (חלקו של AV1 בקטלוג של YouTube): https://www.tvtechnology.com/platform/streamin g/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 (פעילות מסחרית של DTV+, יוני 2026): https://mpegh.com/brazil/ ; CartaCapital, על VVC ב־: https://www.cartacapital.com.br/?p=628337 ; Olhar Digital, על מקלטים בטלוויזיות: https://olhardigital.com.br/2026/05/01/pro/em-breve-a-tv-aberta-vai-mudar- /no-brasil TV 3.0
- 9to5Google, "YouTube adds AI-upscaled 'Super Resolution' for low-quality videos", 29 באוקטובר 2025: https://9to5google.com/2025/10/29/youtube-ad /ds-ai-upscaled-super-resolution-for-low-quality-videos
- Meta, "How Meta Deployed Super Resolution at Scale to Transform Video Quality", @Scale, 16 באוקטובר 2025: https://atscaleconference.com/how-met /a-deployed-super-resolution-at-scale-to-transform-video-quality
- Samsung, 8K TV highlights (NQ8 AI Gen3, 768 רשתות נוירונים): https://www.samsung.com/us/tvs/8k-tv/highlights/

צילום ותמונות

- 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 ו־Visionary.ai, מעבד ISP מלא מבוסס AI, 5 בינואר 2026: https://www.edge-ai-vision.com/2026/01/chipsmedia-and-visionary-ai-unveil-the-wo rlds-first-ai-based-full-image-signal-processor-redefining-the-future-of-image-quality ; ריאיון ב־PetaPixel: https://petapixel.com/2026/01/18/a-look-at-t he-worlds-first-full-ai-based-image-signal-processor
- JPEG committee, הודעה לעיתונות מהישיבה ה־103 (השיפורים, הסיבוכיות והמימושים בסמארטפונים של JPEG AI): https://jpeg.org/items/20240507_press.html
- G. J. Sullivan, introduction to ISO/IEC JTC 1/SC 29, אפריל 2026 (מצב JPEG AI; נתוני ECM ו־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
- IEEE 1857.11 working group (קידוד תמונה מבוסס רשתות נוירונים): https://sagroups.ieee.org/fvc/
- Chrome 155 ו־JPEG XL, באוקטובר 2026: https://xenospectrum.com/en/chrome-155-jpeg-xl-rust/

שיחות ופגישות

- Microsoft, "Satin: Microsoft's latest AI-powered audio codec for real-time communications", 17 בפברואר 2021: https://techcommunity.microsoft.com/t 5/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 ביולי 2026: https://techcommunity.microsoft.com/blog/linuxandopen sourceblog/announcing-the-open-source-release-of-ml-video-codec-mlvc/4539875 ; קוד: https://github.com/microsoft/mlvc
- Streaming Learning Center, "Microsoft MLVC: Interoperability, Power Realities, and the Neural Compression Patent Landscape": https://streaminglear ningcenter.com/articles/microsoft_mlvc.html ; ניתוח של iReplay: https://ireplay.tv/blog/mlvc-microsoft-neural-video-codec-open-source-streaming/
- הערות הגרסה של Opus 1.6: https://opus-codec.org/demo/opus-1.6 ; Opus 1.5: https://opus-codec.org/demo/opus-1.5

- Meta Engineering, "MLow: Meta's low bitrate audio codec", 13 ביוני 2024: https://engineering.fb.com/2024/06/13/web/mlow-metas-low-bitrate-audio-codec/
- Microsoft, Copilot+ PCs (NPU של 40 TOPS ומעלה): https://www.microsoft.com/en-us/windows/copilot-plus-pcs ; Apple, הכרזת M4 (Neural Engine של 38 TOPS): https://www.apple.com/newsroom/2024/05/apple-introduces-m4-chip/

בינה מלאכותית יוצרת

- Kyutai, כרטיס המודל של Mimi (12.5 Hz, 1.1 kbps): https://huggingface.co/kyutai/mimi
- Lightricks, "LTX-Video: Realtime Video Latent Diffusion" (VAE לווידאו ביחס 1:192): https://arxiv.org/abs/2501.00103
- הודעה לעיתונות מהישיבה ה־153 של MPEG (Video Coding for Machines בשלב DIS): https://www.mpeg.org/wp-content/uploads/2026/03/press_release_153.pdf

משחקים וגרפיקה

- NVIDIA, אימוץ DLSS 4: ההשקה ב־Computex 2025 (75 ויותר מ־125): https://www.nvidia.com/en-us/geforce/news/125-dlss-4-multi-frame-gen-games-more-announced-computex-2025/ ; Gamescom 2025 (יותר מ־175): https://www.nvidia.com/en-us/geforce/news/gamescom-2025-dlss-4-full-ray-tracing-game-announcements/ ; CES 2026 (יותר מ־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", פברואר 2026: https://www.ubisoft.com/en-us/studio/laforge/news/415slmB3ZGzv8d2LhO9QT/shipping-neural-texture-compression-in-assassins-creed-mirage
- TweakTown על הדגמת הטקסטורות הנוירוניות של NVIDIA ב־GTC 2026 (מ־6.5 GB ל־970 MB): 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 retail, 26 בפברואר 2026: https://devblogs.microsoft.com/directx/shader-model-6-9-retail-and-more/ ; על התוכניות ל־GDC 2026: https://devblogs.microsoft.com/directx/shader-model-6-9-and-the-future-of-cooperative-vector/ ; cooperative vector: https://windowsreport.com/microsoft-showcases-neural-rendering-and-dxr-2-0-plans-for-directx-at-gdc-2026/

ערוצים ייעודיים

- Global Trade Alert על הודעה מס׳ 49 לשנת 2026 של סין: 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 (הספים לפלטפורמות קרקעיות בלתי מאוישות): https://www.armyrecognition.com/news/army-news/2026/china-tightens-drone-export-controls-raising-supply-chain-risks-for-the-us-europe-and-ukraine
- DroneLife, ה־FCC מוסיפה רחפנים ורכיבים מתוצרת חוץ ל־Covered List, 22 בדצמבר 2025: https://dronelife.com/2025/12/22/fcc-adds-foreign-made-drones-and-components-to-covered-list-citing-national-security-risks/

סיליקון

- 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 ; דף השבב של MIT (4K30 ב־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 ; חומר משלים (סיבוכיות לפי מודול): https://openaccess.thecvf.com/content/CVPR2025/supplemental/Jia_Towards_Practical_Real-Time_CVPR_2025_supplemental.pdf
- NNVC וספריית ההסקה במספרים שלמים SADL: https://arxiv.org/abs/2309.05846
- Tencent Canghai V2, 27 במאי 2026 (Shanghai Securities News דרך 10jqka): https://stock.10jqka.com.cn/20260527/c677016020.shtml ; סיכום היכולות: https://www.besthub.dev/articles/tencent-s-canghai-v2-chip-enters-mass-production-what-the-new-capabilities-mean-for-video-encoding-bcfc18dbfa29
- ה־IP של Chips&Media למפענח AV2, 12 במאי 2026: https://www.design-reuse.com/news/202530501-chips-media-completes-development-of-next-gen-av2-hw-decoder-ip/ ; רישיון AV2, 17 ביוני 2026: https://www.edge-ai-vision.com/?p=58270
- Allegro DVT, IP של מפענח AV2, 16 ביוני 2026: https://www.edge-ai-vision.com/2026/06/allegro-dvts-pulsar-decoder-ip-adds-support-for-av2-video-code/
- VeriSilicon, VC9800D עם AV2, 10 ביוני 2026: https://www.businesswirechina.com/en/news/63176.html

תקנים ופטנטים

- הודעה לעיתונות מהישיבה ה־155 של MPEG (קול קורא משותף להצעות מעבר ל־VVC), 5 באוגוסט 2026: https://www.mpeg.org/wp-content/uploads/2026/08/ag3w254SN26474Press-Release-MPEG155.pdf
- הודעה לעיתונות מהישיבה ה־152 של MPEG (תוצאות הקול הקורא לראיות; המהדורות של VVC ו־VSEI), 30 באוקטובר 2025: https://www.mpeg.org/wp-content/uploads/2025/11/Press-Release-MPEG152.pdf ; טור MPEG ב־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, מיפוי הפטנטים שמעבר ל־VVC (H.267): https://greyb.com/blog/exclusive-video-codec-landscape/ ; על רישוי AV2 ועל מכירת תיק הפטנטים של Alibaba: https://greyb.com/blog/av2-licensing/

- MLex, "Fragmented codec licensing exposes need for clearer rules for Chinese tech majors", 27 במרץ 2026: https://www.mlex.com/mlex/articles/2458402/fragmented-codec-licensing-exposes-need-for-clearer-rules-for-chinese-tech-majors
- Access Advance, תעריפי התמלוגים של MCBA ו־VVC Advance נכון ל־1 ביולי 2026: https://accessadvance.com/wp-content/uploads/2025/07/MCBA-Program-Overview-September-2026.pdf

כסף וחברות

- InterDigital רוכשת את Deep Render, 30 באוקטובר 2025: https://www.advanced-television.com/2025/10/30/interdigital-acquires-deep-render
- TechCrunch, "Apple acquired a startup using AI to compress videos", 27 במרץ 2023: https://techcrunch.com/2023/03/27/apple-acquired-a-startup-using-ai-to-compress-videos
- Neowin, "Sony acquires AI video startup iSIZE", נובמבר 2023: https://www.neowin.net/news/sony-acquires-ai-video-startup-isize-to-enhance-playstation-streaming/

דחיסה נלמדת · ניגון × סיליקון

קידוד נוירוני ב־2026

בשטח הוא פועל סביב הקודק, לא במקומו

לכל מספר מרכזי בדוח הזה מצוינים המקור והתאריך. מספר שהמחבר גזר בעצמו מסומן כאומדן; מספר שרק המוכר מדווח עליו מסומן כטענת החברה.

תאן שואי

טכנולוג ומייסד · חוקר ומפתח בינה מלאכותית

tanshuai.com · i@tanshuai.com

מהדורה פומבית · הנתונים נכונים ל־12 בנובמבר 2026