הסטודיו של מל | 20+ שנות ניסיון בפיתוח וקידום אתרים

איך לחסוך ב-usage בקלוד קוד

חמש פעולות שעובדות אצלי

חמש פעולות שאני עושה בעבודה היומיומית, למה הן כנראה עוזרות ומה מהן נמדד ומה עוד לא. אצלי זה עובד. כדאי לבדוק מה מתאים לך.

הדרכת וידאו מעשית: איך לחסוך ב-usage בקלוד קוד

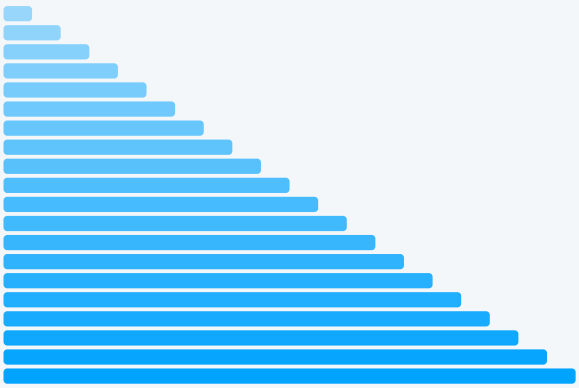
צפייה מודרכת בליווי צעד אחר צעד מאת נורית מל מאטיס

מעדיפים לראות הכל בעיניים? סרטון הדרכה מצולם המלווה את חמש הפעולות צעד אחר צעד.

הסרטון ביוטיוב: https://youtu.be/y5WQtoO_PFg

1. למה שיחה ארוכה עם קלוד קוד עולה יותר?

Claude לא זוכר שיחה. בכל הודעה נשלחת אליו כל השיחה מההתחלה והוא קורא אותה מחדש. לכן ככל שהשיחה מתארכת יש לו יותר לקרוא בכל הודעה. כנראה שגם זה אוכל יותר מהמכסה.



210
יחידות קריאה בשיחה אחת ארוכה

110
יחידות קריאה בשתי שיחות קצרות, חצי מההודעות בכל אחת

דוגמה מופשטת: כל הודעה היא יחידה אחת, בלי המטמון ובלי הטעינה ההתחלתית של שיחה חדשה. היא מראה את הכיוון ולא את החיסכון בפועל. כמה בדיוק זה משפיע על המכסה לא נמדד.

2. מהן חמש הפעולות שחוסכות ב-usage בקלוד קוד?

נמדד יש נתון **מאומת** נבדק שזה נכון **סביר** הגיוני, לא נמדד

לא נבדק רק מתחילים לבדוק

1 שולטים על ה-Context

נמדד —

מה עושים

כששיחה מגיעה ל-30 אחוז מחלון הקונטקסט מסכמים אותה ופותחים שיחה חדשה. בעבודה כבדה מחכים לסוף היחידה הנוכחית, עד 45 אחוז.

איך עושים בקלוד קוד

1. בודקים כמה מהחלון מלא (הפקודה `/context` מראה את הפירוט).
2. ב-30 אחוז מבקשים מ-Claude סיכום קצר: מה נעשה, איפה עצרנו ומה הצעד הבא.
3. פותחים שיחה חדשה בתיקייה ומדביקים את הסיכום.

למה זה עוזר

שיחה קצרה אומרת פחות טקסט לקריאה בכל הודעה. שיחה חדשה אינה ריקה לגמרי: אצלי היא מתחילה על כ-9 עד 10 אחוז, בגלל קבצי ההוראות, הזיכרון והכלים.

מה לא הוכח

כמה זה חוסך ממכסת 5 השעות ומהמכסה השבועית. חלק מהקריאה החוזרת זול יותר בגלל מטמון ולא ידוע איך המכסה סופרת אותו.

2 בוחרים מודל לפי המשימה

סביר —

מה עושים

Haiku למשימה מכנית, Sonnet לעבודה רגילה ו-Opus למשימת עומק. אני מתחילה בכל שיחה ב-Sonnet ועוברת ל-Opus רק כשצריך.

איך עושים בקלוד קוד

1. פותחים את בורר המודלים (אפשר גם עם `/model`).
2. בוחרים לפי המשימה: ניסוח או סידור פשוט מול ניתוח עמוק.

למה זה עוזר

דרך כלל, מודל חזק יותר אומר יותר זמן ויותר מהמכסה ולא כל משימה צריכה אותו.

מה לא הוכח

השימוש ב-Haiku והחיסכון מההחלפה בין מודלים עוד לא נמדדו אצלי.

3 מגדירים רמת מאמץ לפי המשימה

לא נבדק —

מה עושים

רמת המאמץ קובעת כמה Claude "חושב" לפני שהוא עונה. Low למשימה פשוטה, High למשימה מורכבת.

איך עושים בקלוד קוד

1. מוצאים את בורר רמת המאמץ ליד בורר המודל.
2. למשימה פשוטה מורידים. למשימה מורכבת מעלים.

למה זה עוזר

יותר מאמץ אומר יותר זמן וכנראה יותר מהמכסה ולכן לא כדאי להשאיר אותו גבוה כל הזמן.

מה לא הוכח

הכול. אני רק מתחילה לבדוק את זה ולכן אין לי עדיין נתון.

4 פותחים כל שיחה בתיקייה של הנושא

סביר —

מה עושים

תיקייה אחת לכל נושא או פרויקט ושיחה אחת לכל נושא. בתיקייה שומרים קובץ הוראות קצר (`CLAUDE.md`) וכך Claude יודע מה הנושא, איפה הקבצים ומה חשוב כבר מההודעה הראשונה.

איך עושים בקלוד קוד

1. פותחים את השיחה בתיקייה של הנושא.
2. הקובץ לא נוצר לבד. כותבים אותו בעצמנו, או מריצים בתיקייה את `/init` ו-Claude כותב אותו אחרי שהוא סורק אותה.
3. שומרים אותו קצר: רק מה שמיוחד לתיקייה הזאת.

למה זה עוזר

השיחה טוענת מיד את ההוראות והקבצים של הנושא, בלי לחפש ובלי לקרוא קבצים שלא קשורים. בדרך כלל זה אומר עבודה מהירה יותר ופחות קונטקסט.

מה לא הוכח

הקובץ לא חוסך בעצם קיומו. הוא חוסך רק כשהוא קצר וכל הוראה נמצאת במקום שבו היא נחוצה. כמה זה חוסך בפועל לא נמדד.

5 מצמצמים צילומי מסך ו-PDF

מאומת שהם נשארים —

מה עושים

צילום מסך או קובץ PDF שנכנסו לשיחה נשארים בה, ו-Claude קורא אותם שוב בכל הודעה.

איך עושים בקלוד קוד

1. מעתיקים טקסט במקום לצלם.
2. חותכים את התמונה לחלק שצריך.
3. אחרי כמה תמונות פותחים שיחה חדשה.

למה זה עוזר

יותר קבצים בשיחה אומרים שיחה יקרה יותר בכל הודעה שאחריהם.

מה לא הוכח

כמה בדיוק זה חוסך. את העובדה שהקבצים נשארים בשיחה בדקתי ואת גודל החיסכון לא.

3. מה המדידה מראה על החיסכון ב-usage?

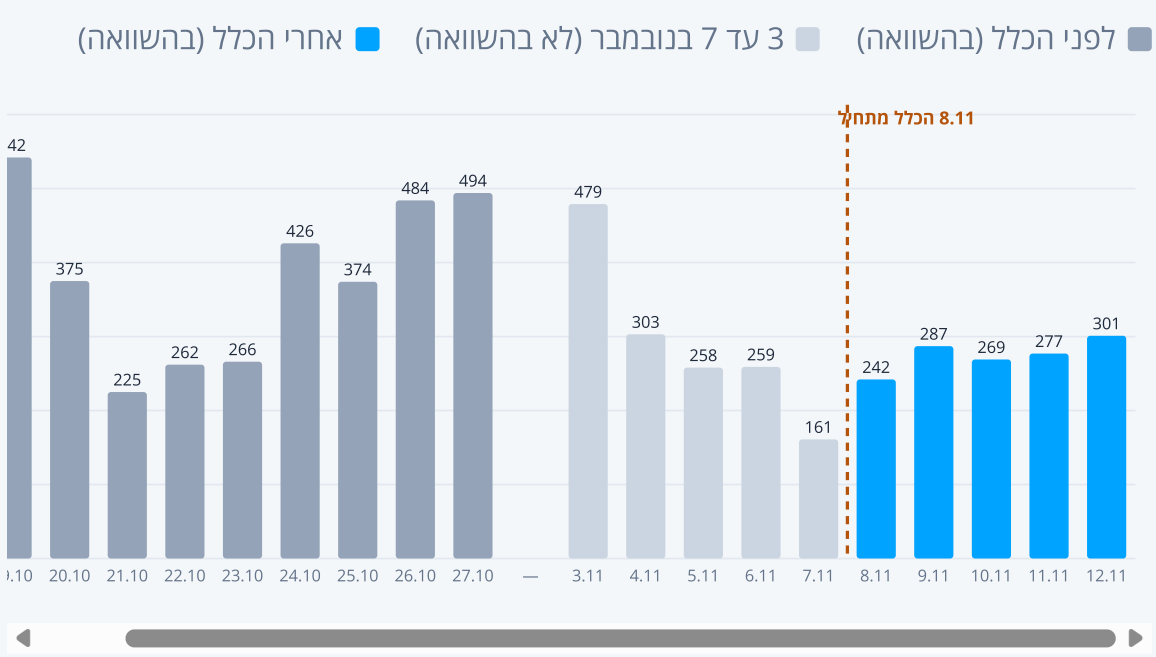
ב-8 בנובמבר 2026 הוספתי כלל: שיחה שמגיעה ל-30 אחוז עוברת לשיחה חדשה. מדדתי כמה קונטקסט Claude קורא בממוצע בכל הודעה, מתוך התמלולים של השיחות במחשב שלי.

354
אלפי טוקנים להודעה, 19 עד 27 באוקטובר (לפני)

270
אלפי טוקנים להודעה, 8 עד 12 בנובמבר (אחרי)

24%

ירידה בממוצע, בלי ירידה במספר ההודעות ביום



ממוצע הקונטקסט שנקרא בכל הודעה, באלפי טוקנים, לפי יום. 28 באוקטובר עד 2 בנובמבר: אין נתונים בתמלולים במחשב וזה לא אומר שלא עבדתי. 12 בנובמבר הוא יום חלקי. הממוצעים בהשוואה משוקללים לפי מספר ההודעות.

הסתייגויות

- המדד הוא קירוב לעלות ולא מונה רשמי של Claude.
- המדידה כוללת רק שיחות שנשמרו במחשב אחד, בלי הפרדה בין חשבונות.
- זה שבוע מול שבוע ולא ניסוי מבוקר. גם סוג העבודה השתנה בין השבועות.
- הנתונים מראים שהקונטקסט הממוצע ירד. את השיפור במכסה אני מרגישה והוא החוויה שלי ולא נתון רשמי.

4. איך אני עובדת

לפני הכלל הייתי מגיעה למכסת 5 השעות בקלות ולפעמים גם למכסה השבועית. אז הייתי צריכה לחכות יום או יומיים.

אחרי שהוספתי את הכלל זה השתפר אצלי באופן ניכר. זו החוויה שלי ולכן אני כותבת אותה כך ולא כהבטחה.

היום כל נושא יושב בתיקייה שלו עם קובץ הוראות קצר. שיחה שמגיעה ל-30 אחוז מסכמת את עצמה ועוברת לשיחה חדשה שממשיכה מיד מהצעד הבא. את ההעברה האוטומטית עצמה עוד לא בדקתי מקצה לקצה בשיחה אמיתית ואעדכן כשאבדוק.

5. הדרכה אישית למתחילים

כ-40 דקות על איך לחסוך ב-usage בקלוד קוד, בצעדים שמתאימים לעבודה שלך. לפרטים ולתיאום פונים אליי בוואטסאפ.

054-896-4830

דברו איתי בוואטסאפ!

אני נורית מל מאטיס ואני איתך, לא מעליך. אני מלווה אותך עד שהפרויקט עובד עבורך: מקשיבה למה שהעסק שלך באמת צריך ומסבירה מה לדעתי כדאי לעשות ולמה. כל פרויקט עובר דרכי אישית מהשיחה הראשונה ואני כאן גם אחרי שהפרויקט יצא לדרך. כשנדרש בעל מקצוע מתחום אחר אני ממליצה על מקצוענים בתחומם ומשתפת איתם פעולה כדי להרכיב יחד את הפאזל השלם. 20+ שנות ניסיון ו-500+ אתרים.

נכתב ונערך באהבה | הסטודיו של מל

מעבר לאתר הסטודיו: **nuritmellmatis.co.il**