דעה

הגבולות של מודלי LLM

PhotoCraft, עורך תמונות מוקדם שנבנה ב-AI, REA, ערכת כלים להנדסה לאחור,
ופורט חדש של משחק רטרו כמעט בכל שבוע שכנעו רבים ש-Adobe מתה
ושהמתכנתים הבאים בתור. קראתי את הקוד ואת המספרים.

/Fabio Akita • ailair.akitaonrails.com/he/limits

> ***מודלי LLM יכולים לשכפל את מה שפתוח, אבל הם לא יכולים לנחש את מה שפרטי.***
>
> *LLM consegue replicar o que é aberto, mas não adivinha o que é privado.*

1. במה הוא טוב

להעתיק את מה שכבר ציבורי

המודלים למדו מקוד ציבורי. כשמשהו דומה כבר פורסם, סוכן בונה אותו מחדש מהר ובזול.

81.5%

מהתרומות ב-GitHub קורות במאגרים פרטיים (Octoverse 2025)

TB 67.5

מהקוד הציבורי ב-The Stack v2, סוג הנתונים שמודלים מתאמנים עליו

להעתיק את מה שכבר ציבורי

- בגלל זה אדם אחד יכול היום להוציא קליינט נייטיב ל-Spotify או ל-WhatsApp ב-Rust. השירות, הספריות והדרך כבר היו קיימים. ללכת שוב בדרך מוכרת נהיה זול.
- איפה שאין שום דבר דומה, הסוכן חוזר לניסוי וטעייה. הוא צריך חומרי ייחוס שלא פורסמו, בדיקות ובדיקה אובייקטיבית, שנקראת לעיתים אורקל, כדי לדעת אם התוצאה נכונה.
- הקוד שמשלם משכורות הוא ברובו פרטי. מודל כללי לא אומן עליו כברירת מחדל. הוא יכול לנתח קוד שתספקו דרך כלי מאושר, עם הקונטקסט ובקרות הגישה המתאימים.



2. הנדסה לאחור

פירוק בינאריים קיים כבר 40 שנה

כל אמולטור שאי פעם השתמשתם בו קיים כי מישהו פירק בינארי. ה-AI הצטרף לפס ייצור שכבר עבד.

פירוק בינאריים קיים כבר 40 שנה

- בתי המשפט בארה"ב קבעו את היסודות ב-1992 (Sega v. Accolade) וב-2000 (Sony v. Connectix): פירוק קוד כדי ללמוד איך הוא עובד הוא שימוש הוגן (fair use).
- דקומפיילרים מבוססי LLM כותבים קוד קריא יותר מהמסחריים, עם נכונות פונקציונלית נמוכה ב-52% (DecompileBench, 2025). העוזרים שבשימוש אמיתי נותנים שמות למשתנים על גבי מה ש-Ghidra או IDA הפיקו.
- הכלים בשלים, ורבים מהם חינמיים. ה-NSA שחררה את Ghidra ב-2019, באותה שנה שבה ה-decomp של Super Mario 64 יצא לציבור.
- REA, עם קרוב ל-200,000 כוכבים ב-GitHub, נותן את אותם כלים לסוכן שלכם. צריך את הבינארי, את הכלים מותקנים והרבה טוקנים: בדיקת ה-SQLite שלו עצמו לקחה 42 קריאות לכלים ו-300,000 טוקנים של קלט. התיעוד שלו אומר לא להגיד אף פעם שהוא משכפל תוכנה.



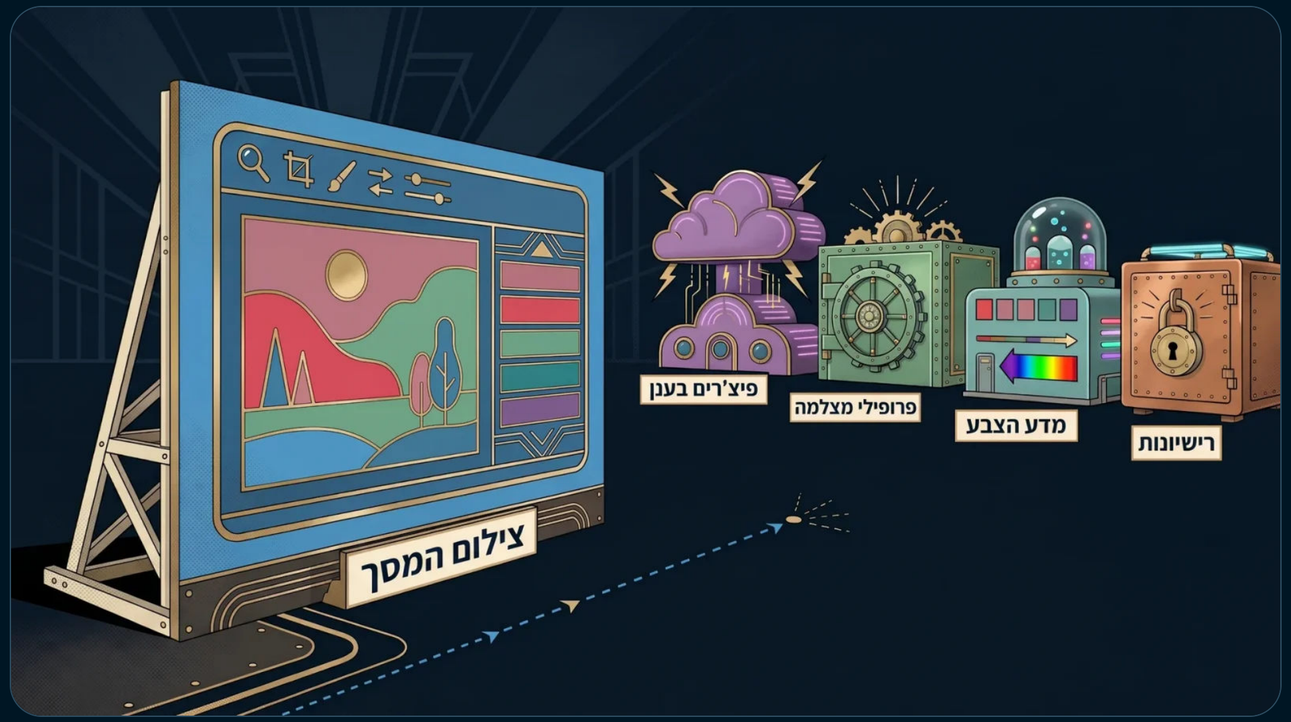
3. משחקים ישנים

למה משחקים ישנים נופלים ראשונים

פורט רטרו חדש יוצא כמעט כל יום. למשחקים ישנים יש כל יתרון שחסר למוצר חי.

למה משחקים ישנים נופלים ראשונים

- הבינארי זעיר. Super Mario 64 שוקל 8 MB. Baldur's Gate 3 שוקל 150 GB.
- PS4, PS5, Xbox ו-Switch הם מחשבים או טלפונים בקופסה אחרת. המעבד שלכם כבר מריץ את הקוד שלהם, ו-Wine, DXVK ו-Proton פתרו את רוב תרגום הגרפיקה לפני שנים.
- החומרה מופתה פקודה אחר פקודה במשך 25 שנה, והיעד אף פעם לא משתנה. אף אחד לא מעדכן את Zelda של 1998.
- ה-AI האיץ את העבודה השחורה. המפתחים של ה-recomps של Zelda ושל Banjo קוראים לפורטים שסוכנים הכינו "slopcomps": לא נבדקו, וטולאו בצורה שמקשה לתחזק אותם.



◆ 4. השיבוטים של Adobe

צילום מסך של Photoshop

PhotoCraft נראה כמו Photoshop בצילום מסך. מה שאנשים משלמים עליו ל-Adobe נמצא במקום אחר.

צילום מסך של Photoshop

968

issues פתוחים ב-PhotoCraft, נבדק ב-13 באוקטובר 2026

- PhotoCraft נכתב ב-Rust עם Claude. ה-README שלו עצמו אומר early alpha. תמונות מעל 8,192 פיקסלים חוזרות ל-CPU, ניהול הצבע בסיסי, וכ-100 issues פתוחים עוסקים בקובצי PSD.
- Camera Raw מביא איתו 20 שנה של פרופילים לכל מצלמה ועדשה. לשיבוט או שיש את הנתונים האלה, או שאין.
- רוב התוכנה היום רצה כשירות. הקוד שמאחורי Google Docs אף פעם לא יוצא מהשרתים של Google, אז אין בכלל בינארי.

13

ימים אחרי שהמאגר נוצר

- Generative Fill וכמה מה-Neural Filters רצים בשרתים של Adobe. הקוד הזה אף פעם לא מגיע למחשב שלכם, אז אין מה לפרק.
- צבע לדפוס הוא תחום בפני עצמו. GIMP נולד ב-1995 ועדיין אין בו עריכת CMYK בגרסה 3.0. צבעי Pantone דורשים רישיון, ואפילו Adobe איבדה אותם ב-2022.

> *PhotoCraft היום הוא כמו רפליקה של פרארי שבנויה על חיפושית של פולקסווגן: למרכב יש את הצורה הנכונה, מצלמים אותו ורוב האנשים אומרים "וואו, פראררי". ואז מסובבים את המפתח ויוצא המנוע של החיפושית.*
>
> *O PhotoCraft hoje é aquela réplica de Ferrari montada em cima de um Fusca: a carroceria tem a forma certa, você tira .uma foto e a maioria das pessoas diz "uau, uma Ferrari". Aí você liga e sai o motor de Fusca*



◆ 5. העלות האמיתית

כמה עולה שכתוב גדול

המספר הציבורי הטוב ביותר מגיע מ-Bun, סביבת הריצה של JavaScript, ששוכתבה מ-Zig ל-Rust עם סוכנים במאי 2026.

כמה עולה שכתוב גדול

$ 165K

לפי מחירי API

11

ימים, 535,000 שורות Zig

64

סוכני Claude בשיא

- זה עבד בזכות ארבעה דברים: היוצר של Bun עצמו ליד ההגה, קוד המקור המלא והעובד, כ-60,000 בדיקות לכל פלטפורמה לבדוק מולן כל שינוי, ו-Anthropic, שמחזיקה ב-Bun, ששילמה.
- לשיבוט של Photoshop אין אף אחד מהארבעה. אין קוד מקור להעביר, אין חבילת בדיקות שאומרת "תואם ל-Photoshop", ואין יוצר. ו-Photoshop הוא מוצר בן 36, גדול בהרבה מסביבת ריצה.
- הקומפיילר ל-C של Anthropic, פברואר 2026: 16 סוכנים, פחות מ-20,000 דולר, והוא מקמפל את Linux. הוא עדיין שואל את ה-assembler וה-linker של GCC, והקוד שהוא מייצר רץ לאט יותר משל GCC עם כל האופטימיזציות כבויות.

6. המחיר

“בעוד שנתיים זה יהיה בחינם”

המחיר של המודלים הטובים ביותר כמעט לא זז מאז 2023. מה שירד מהר הוא המחיר של רמת איכות שכבר ישנה.

$ 60

למיליון טוקנים של פלט, GPT-4, מרץ 2023

$ 50

למיליון טוקנים של פלט, המודלים הטובים ביותר, אוקטובר 2026

“בעוד שנתיים זה יהיה בחינם”

- החזית נהייתה זולה פי 1.2 עד פי 3 בשלוש שנים וחצי, תלוי אם מדובר בטוקנים של קלט או של פלט.
- תעשו את החשבון על Bun: 165,000 דולר חלקי 100 זה 1,650 דולר, לפורט עם קוד המקור, הבדיקות והיוצר ביד. זול וטריוויאלי הם דברים שונים.
- רמת איכות קבועה נהיית זולה בערך פי 5 עד פי 10 בכל שנה, לפי ההערכה של Epoch AI, והם צופים שזה יאט.
- המודלים עצמם התכנסו, והשיפורים האחרונים באו מכלי הסוכנים שסביבם. אני מסביר את זה בעמוד על AGI.

המתכנתים מתים?

לכל שכתוב ולכל decomp שעובדים יש מהנדס ליד ההגה.

15%+

במודעות הדרושים לתוכנה בארה"ב מאז ש-Claude Code יצא בפברואר 2025, בזמן שכלל המודעות ירדו ב-7% (Indeed, יולי 2026)

71%

מהצמיחה הזו הגיעו ממשרות סניור

המתכנתים מתים?

- מודעות הדרושים לתוכנה עדיין נמוכות בכ-27.5% מפברואר 2020. פחות אנשים מתקבלים לעבודה כדי להקליד. מי שיודע לשפוט את התוצר מתקבל.
- בבנצ'מרק שלי על 7 שפות, לתת למודלים החלשים בדיקות שינה את התוצאות שלהם יותר מכל שפה.
- לשכתב את Bun דרש את היוצר של decomps-ב. שעובדים, מי שמכוון את הסוכן הם אנשים שמכירים את החומרה.
- הנדסה לאחור נהייתה זולה ונגישה יותר, כמו כש-Ghidra יצא ב-2019. זו עדיין עבודה של מהנדס.

> ***כותרות של "נגמר" מביאות קליקים ולא מתקמפלות.***
>
> *.Manchete de "acabou" rende clique e não compila*

המסקנות

1

AI מעתיק מהר את מה שציבורי. רוב הקוד שחשוב הוא פרטי.

2

הנדסה לאחור קיימת כבר עשרות שנים. סוכנים האיצו את העבודה השחורה שלה.

3

משחקים ישנים נופלים ראשונים: הם קטנים, ממופים כבר עשרות שנים, ולא משתנים אף פעם.

4

צילום מסך של שיבוט זה קל. שירותי הענן, הנתונים והרישיונות לא נמצאים באף בינארי.

5

השכתוב של Bun דרש את היוצר שלו, את קוד המקור שלו, את הבדיקות שלו ו-165,000 דולר.

6

המודלים הטובים ביותר כמעט לא הוזלו, והם התכנסו.

7

העבודה שגדלה היא שיקול דעת: בדיקות, בדיקות אובייקטיביות שנקראות אורקלים ו-code review.

כותרות של “נגמר” מביאות קליקים ולא מתקמפלות.

Manchete de "acabou" rende clique e não compila.

לקריאה נוספת: הפוסטים שלי והמקורות

להעתיק את מה שכבר ציבורי

- המגבלות של מודלי LLM: טובים בשחזור מה שקיים, יקרים במה שלא ↗
- GitHub Octoverse 2025: 81.5% פרטי ↗
- The Stack v2: מאגר הנתונים של הקוד הציבורי ↗

פירוק בינאריים קיים כבר 40 שנה

- הנדסה לאחור עם AI: האם Adobe מתה? ↗
- Phoenix וה-PC BIOS בשיטת clean room ↗
- Sega v. Accolade (1992) ↗
- Sony v. Connectix (2000) ↗
- Ghidra, מבית ה-NSA ↗
- DecompileBench: מודלי LLM מול דקומפיילרים מסחריים ↗
- REA ב-GitHub ↗
- REA: כמה עולה ניתוח אחד ↗

למה משחקים ישנים נופלים ראשונים

- הנדסה לאחור עם AI: האם Adobe מתה? ↗
- איך DriveClub ו-shadPS4 כמעט ניצחו את ה-AI ואותי: איך ללמוד ↗
- Zelda 64: Recompiled ↗
- shadPS4: אמולטור PS4 ↗
- Kotaku: המריבה על ה-“slopcomps” ↗

צילום מסך של Photoshop

- הנדסה לאחור עם AI: האם Adobe מתה? ↗
- PhotoCraft ב-GitHub ↗
- Adobe: ה-Neural Filters צריכים את הענן ↗
- Adobe: פרופילי עדשות ב-Camera Raw ↗
- מפת הדרכים של GIMP: CMYK עדיין בעבודה ↗
- PetaPixel: Pantone עוזבת את Photoshop ↗

כמה עולה שכתוב גדול

- הנדסה לאחור עם AI: האם Adobe מתה? ↗
- Bun: השכתוב ל-Rust, עם מספרים ↗
- Anthropic: קומפיילר C שסוכנים בנו ↗

לקריאה נוספת: הפוסטים שלי והמקורות

"בעוד שנתיים זה יהיה בחינם"

- LLMs + סוכנים = מספיק מוכשרים ↗
- הנדסה לאחור עם AI: האם Adobe מתה? ↗
- Anthropic: תמחור API ↗
- OpenAI: תמחור API ↗
- Epoch AI: מגמות במחירי ה-inference ↗
- Epoch AI: כמה זמן נמשך נטל העלות ↗

המתכנתים מתים?

- יש שפות טובות יותר לתכנות עם LLMs? ↗
- LLMs + סוכנים = מספיק מוכשרים ↗
- הנדסה לאחור עם AI: האם Adobe מתה? ↗
- Indeed Hiring Lab: מודעות דרושים לתוכנה, יולי 2026 ↗

כל נקודה, עם כל הקישורים, באתר

/ailair.akitaonrails.com/he/limits

Fabio Akita • akitaonrails.com • x.com/AkitaOnRails