



הפינה המקצועית של ירדן פלד

TECH עדיף

מהמנוי לשעון: השינוי שיסנה את הדרך שבה סוכנות ביטוח תשלם על AI

בודד יכול לצרוך פי מאה ממשמש רגיל. אף ספק לא יכול לתמחר מנוי אחיד כשהשונות בין הלקוחות היא בסדר גודל כזה.

לכן הכיוון ברור: ככל שהכלים יהפכו לאוטונומיים יותר, כך התמחור יזוז עוד יותר לכיוון מדידת צריכה. סוכנות שתבנה את התוכנית העסקית שלה על ההנחה שהעלות היא סכום חודשי קבוע, תגלה פער בין התקציב למציאות בדיוק ברבעון שבו הפעילות עולה.

למה זה קריטי דווקא בענף שלנו

בענף הביטוח יש עונתיות ברורה, וזה בדיוק מה שהופך את המעבר לצריכה לרגיש כאן יותר מאשר בענפים אחרים.

חשבו על החודשים שבהם מתרכזים חידושי הפוליסות. חשבו על השבועות שאחרי אירוע נזק רחב, כשמחלקת התביעות עובדת בהיקף כפול. חשבו על היום שבו מתפרסם חוזר חדש של רשות שוק ההון וכל הצוות רץ לנתח אותו. בדיוק בשיאים האלה, שבהם ה-AI הכי שימושי, השימוש מזנק, והעלות מזנקת איתו.

שיאי הצריכה בסוכנות ביטוח אינם אקראיים. הם נופלים בדיוק על עונת החידושים, על אירועי נזק ועל פרסום חוזרים.

התוצאה המסוכנת אינה חשבון גבוה. התוצאה המסוכנת היא שסוכנות מגלה את החריגה באמצע השיא, מגבילה את השימוש, ומחזירה את הצוות לעבודה ידנית בדיוק בשבוע שבו הוא הכי עמוס.

הרגע שבו המנוי מפסיק להיות מנוי

סוכן שאני מלווה סיפר לי החודש על שיחה מביכה מול לקוח. באמצע ניתוח סיכונים ללקוח, המערכת הודיעה לו שהמכסה שלו נגמרה. הוא שילם על מנוי חודשי, הוא היה בטוח שהוא קנה שקט תעשייתי, והוא גילה באמצע יום עבודה שהוא קנה בעצם כמות.

זו לא תקלה נקודתית. זו התחלה של שינוי מבני בכל שוק ה-AI, ומי שמנהל סוכנות ביטוח צריך להבין אותו עכשיו, לא בעוד שנה.

מה באמת השתנה

עד לא מזמן, המודל העסקי של כלי ה-AI היה פשוט: תשלום חודשי קבוע, ובתוכו "כמה שצריך". במהלך 2026 המבנה הזה נסדק בכל הספקים הגדולים בו זמנית.

אצל Anthropic, המנוי בתשלום עדיין כולל מכסת שימוש, אבל מעבר לה העבודה ממשיכה בתמחור צריכה לפי תעריפי הטוקנים הרגילים. במקביל נפתחה אפשרות לרכוש מראש חבילות שימוש בהנחה שמגיעה עד 30%, כאשר היתרה משמשת בכל מוצרי החברה ולא רק בצ'אט.

אצל OpenAI התהליך דומה. השינוי המשמעותי ביותר בתמחור השנה היה המעבר מתמחור לפי הודעות לתמחור לפי קרדיטים שנגזרים ישירות מטוקנים. בתוכניות העסקיות, שימוש שחורג מהמכסה הכלולה צורך קרדיטים לפי טבלת תעריפים, וקרדיטים הם גם מה שפותח יכולות מתקדמות כמו מחקר מעמיק, מודלי חשיבה וסוכנים אוטומטיים.

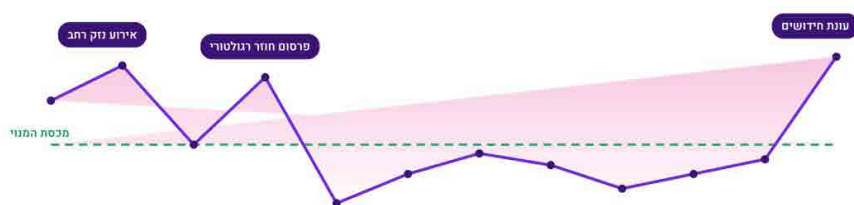
במילים אחרות: המנוי לא נעלם, אבל הוא הפך לרצפה ולא לתקרה. מעליו יש שעון. אצל OpenAI תראו כרגע את כמות השימוש והמנוי הוא חודשי, אבל המגמה ברורה וזה צעד ראשון לקראת מעבר לתשלום לפי שימוש.

למה זה קורה, ולמה זה לא יתהפך

הסיבה אינה חמדנות, אלא חשבון. כל עוד אדם הקליד שאלה וקיבל תשובה, השונות בין משתמשים הייתה סבירה. ברגע שנכנסו סוכנים אוטומטיים, כלים שרצים על קבצים שלמים ותהליכים שפועלים ברקע, אותו מנוי

צריכת AI בסוכנות ביטוח לאורך השנה

המנוי קבוע. הצריכה לא. השיאים מגיעים בדיוק כשהכלי הכי נחוץ.



המשך בעמוד הבא <<<

המשך מהעמוד הקודם <<

דוגמה קטנה: אותה תפוקה, פער של פי ארבעה

נניח סוכנות עם שישה עובדים, שמבצעת כ-600 משימות AI בחודש. 60% מהן מכניות (סיכום שיחה, ניסוח הודעה ללקוח, הגהה), 30% מקצועיות (השוואת כיסויים, ניתוח חוזר, מכתב ערר) ו-10% מורכבות (תיק חריג, תוכנית עבודה). (ראו טבלה)

הפער נוצר משני מרכיבים שמכפילים זה את זה. התאמת המודל למשימה לבדה חוסכת כ-60%, משום שרוב המשימות בסוכנות הן מכניות. ניקוי ההקשר חוסך שוב כמחצית מהנותר, משום ששיחה ארוכה שולחת מחדש את כל ההיסטוריה שלה בכל הודעה.

שימו לב למה שלא מופיע בטבלה: שתי הסוכנויות ביצעו בדיוק את אותה כמות עבודה, באותה איכות. ההבדל כולו בדפוס העבודה, לא בהיקף הפעילות.

שלוש טעויות ניהוליות שיעלו כסף

הראשונה: להתייחס לזה כהוצאת IT. תקציב AI מתנהג כמו תקציב שיווק דיגיטלי, לא כמו רישיון תוכנה. הוא משתנה לפי פעילות, ולכן הוא צריך מדידה חודשית ולא רק אישור שנתי.

השנייה: להשאיר את הבחירה לכל עובד בנפרד. בסוכנות עם עשרה עובדים, ההבדל בין צוות שעובד לפי כללים לצוות שכל אחד בו עושה מה שנוח לו הוא פער של פי כמה בעלות, על אותה תפוקה בדיוק. הפער הזה כמעט תמיד אינו נובע מכמות העבודה אלא מדפוסי עבודה.

השלישית: לקנות עוד מכסה במקום לתקן את דפוס העבודה. רכישת חבילה גדולה יותר מרגישה כמו פתרון, אבל היא רק מאריכה את אותו קצב שריפה. הנחה על מסלול שגוי אינה חיסכון.

חמישה מהלכים להיערכות

1. הפכו את השימוש לנתון מדיד. לפני כל דיון תקציבי, דעו כמה נצרך החודש, על ידי מי ובאילו משימות. מה שלא נמדד לא ינוהל.

2. תרגמו את העלות ליחידת עבודה. לא "כמה עולה לנו AI בחודש", אלא "כמה עולה לנו הפקת מכתב ערר" או "כמה עולה השוואת

פוליסות". ברגע שיש מספר לפעולה, אפשר להשוות אותו לחלופה הידנית ולקבל החלטות עסקיות אמיתיות, כולל תמחור שירות ללקוח.

3. קבעו כללי עבודה כתובים. לא הנחיות כלליות, אלא כללים תפעוליים: אילו משימות מבוצעות באיזו רמת מודל, מתי פותחים שיחה חדשה, איזה אורך תוצר מבקשים, ומה תמיד עובר בדיקה אנושית לפני שהוא יוצא ללקוח או לחברה.

4. נצלו את מבנה התמחור לטובתכם. ההקשר החוזר שנשמר במקום להישלח מחדש זול משמעותית מהקשר שנשלח שוב ושוב, ובחלק מהמודלים ההפרש מגיע לסדר גודל שלם. המשמעות המעשית לסוכנות: מידע קבוע כמו סגנון כתיבה, מוצרים וקהלי יעד צריך לשבת במקום קבוע, ולא להיכתב מחדש בכל פנייה.

5. הגדירו תקרה והתראה. בדיוק כפי שיש תקרת אשראי לספקים, צריכה להיות תקרת צריכה חודשית והתראה כשמתקרבים אליה. סוכנות שמגלה חריגה מהחשבונית איחרה בחודש.

המשמעות העסקית

יש כאן גם הזדמנות, לא רק סיכון. כשעלות הפעולה נמדדת, אפשר לתמחר שירות. סוכנות שיוזעת שניתוח תיק לקוח עולה לה סכום מסוים בזמן ובמשאבי AI, יכולה לבנות סביבו שירות בתשלום במקום לתת אותו כתוספת חינמית. תמחור לפי צריכה מכריח אותנו לעשות משהו שממילא היה צריך לעשות: לדעת כמה עולה כל תהליך בסוכנות.

השורה התחתונה: עידן "שילמתי, מותר לי הכל" מסתיים. מי שיבנה עכשיו כללי עבודה, מדידה ותקציב, יכנס לשלב הבא עם יתרון ברור על פני מי שימתין לחשבונית שתפתיע אותו.

(לחברת הסבר ודוגמאות, כולל סקילים ייעודיים לחיסכון – [לחצו כאן](#))

• הכותב הינו מנהל עדיףTech, בעלים של חברת InsurAI, מומחה להטמעת בינה מלאכותית ואוטומציה בענף הביטוח

מקורות: מרכז העזרה של Anthropic בנושא חבילות שימוש וקרידיטים; מרכז העזרה ודפי התמחור של OpenAI בנושא קרידיטים, תעריפי טוקנים ותמחור גמיש; דפי תמחור ה-API של שני הספקים. הנתונים נבדקו ביולי 2026 והם משתנים מעת לעת. (את המקורות בפונט קטן יותר)

סוכנות א'	סוכנות ב'	
מודל ברירת המחדל	לפי סוג המשימה	החזק ביותר, לכל משימה
אורך שיחה ממוצע	שיחה חדשה לכל נושא	כשלושים הודעות באותו נושא
אורך התוצר	מוגדר בכל בקשה	לא מוגדר מראש
צריכה חודשית יחסית	כ-25%	100%
התוצאה בפועל	נשארת בתוך המנוי	חריגה מהמכסה באמצע החודש ותשלום תוספת
* אלו מספרי המחשה המבוססים על תמהיל משימות טיפוסי ועל יחסי התמחור המפורסמים בין דרגות המודלים. החליפו את ההנחות בנתוני הסוכנות שלכם, והמספר יישאר באותו סדר גודל		