

28/7/2025

הלה צביון אלוני ימין רן משלום

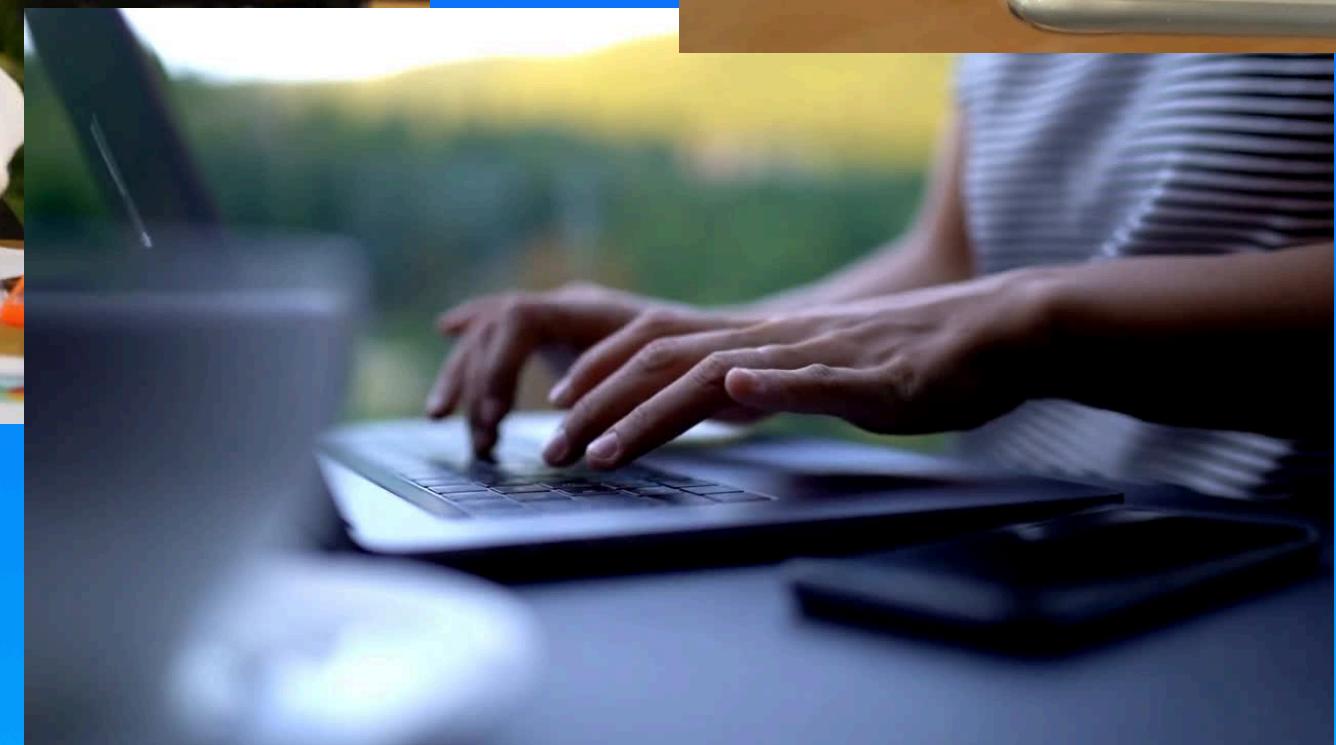
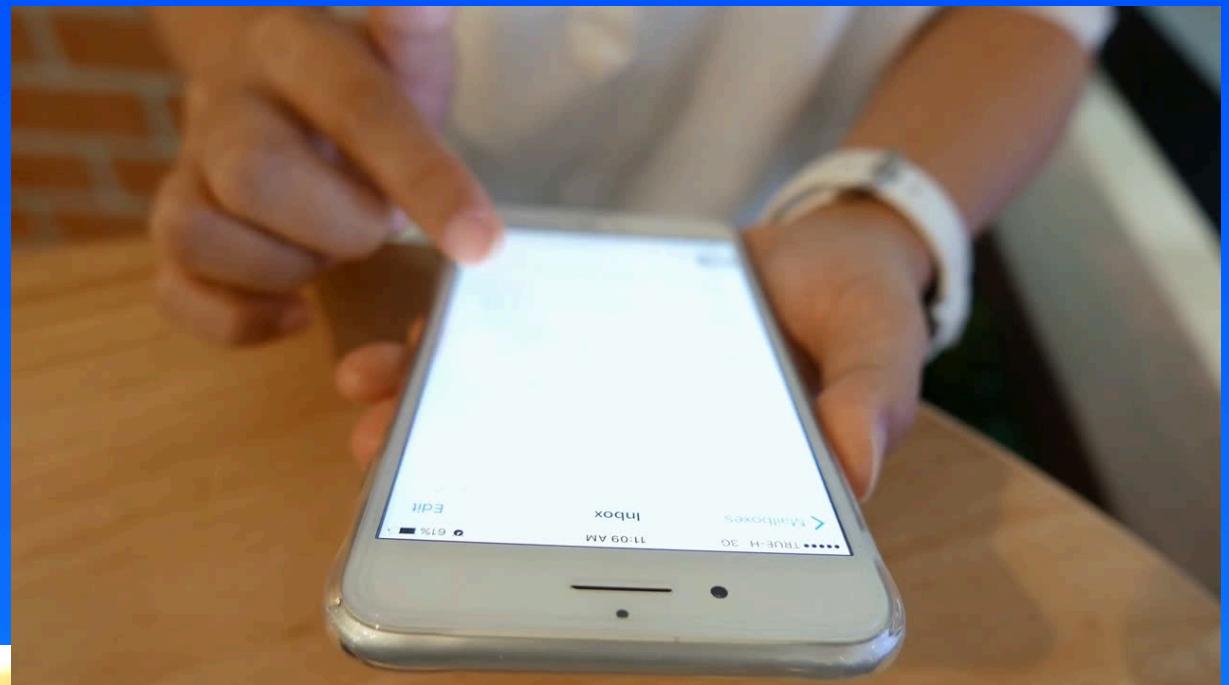
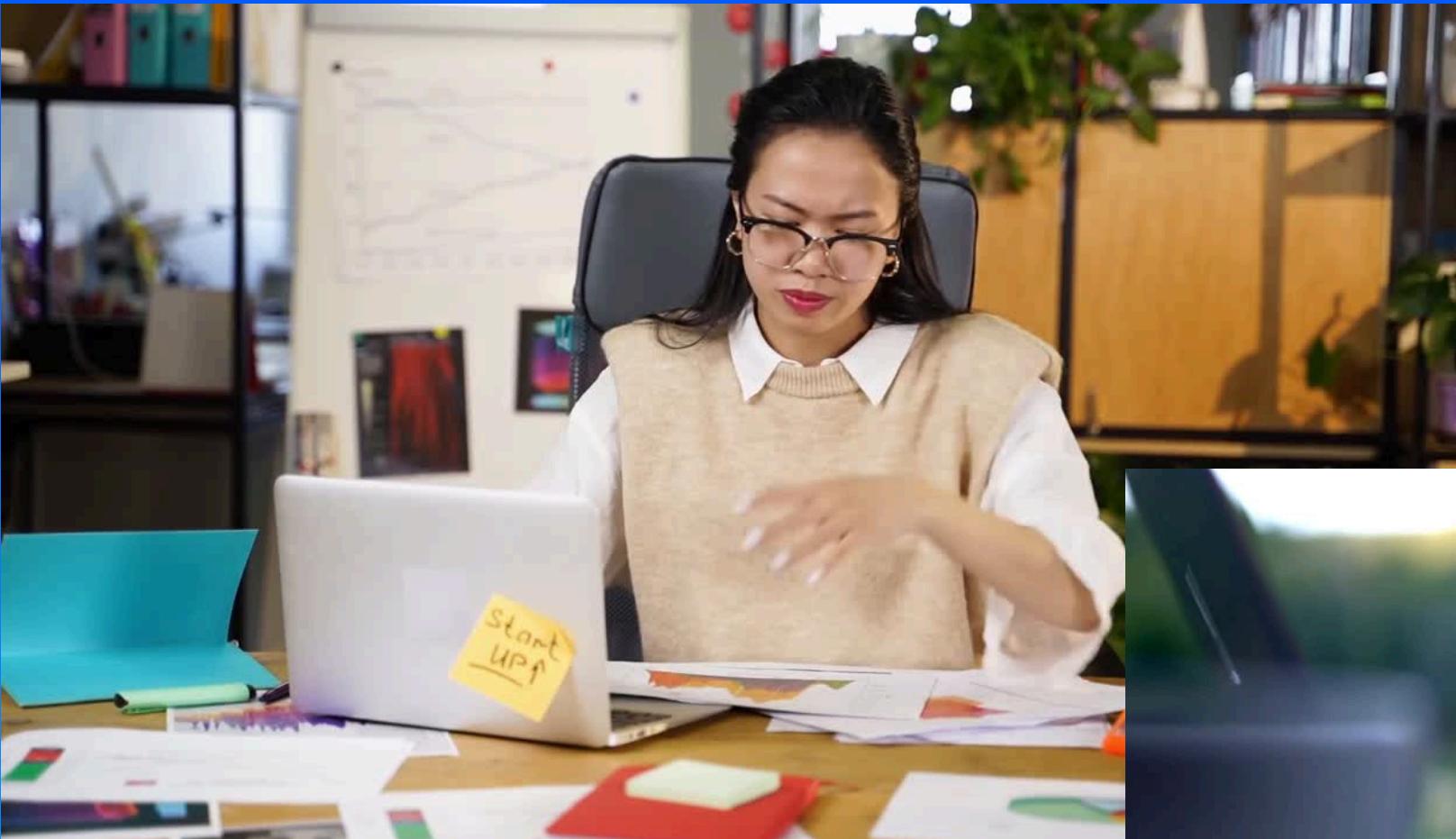
פרויקט גמר מדעי המחשב

# Falcon

Take Your Inbox to New Heights

מנחה: בני בורנפולד

# הצפת מיילים ופגיעה בפרודוקטיביות



# כמה זמן אבו משקיעים במילויים?

**מחקרים מראים כי:**

עובדיו משרדי מבלים עד **8.8 שעות בשבוע** בתעסוקות עם מילויים.

**61%** מהמשתמשים קוראים מילויים דרך **הסмарטפון**

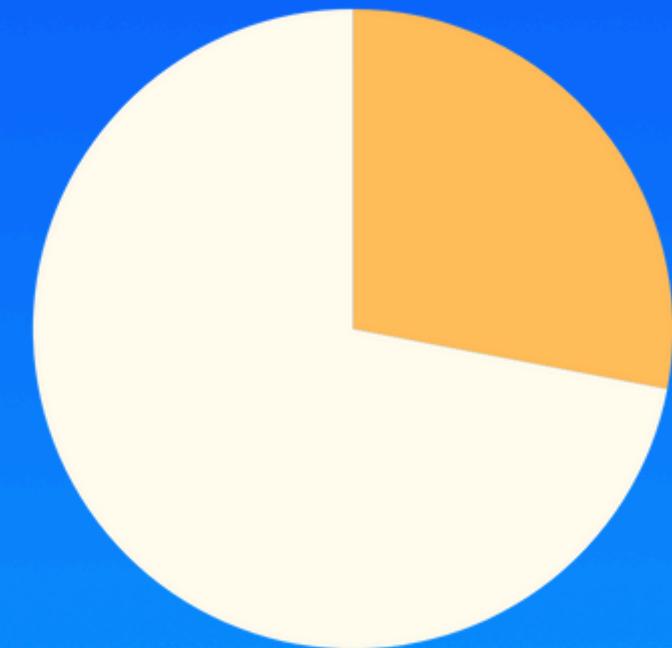
עובדים מבלים ממוצע כ-**28%** זמן העבודה שלהם בטיפול במילויים.



According to a study from **Microsoft**



מקור: **ZeroBounce**, 2024



מקור: **McKinsey Global Institute Analysis**

# במי הבעיה פוגעת בעיקר?

## אנשים מקצוע בעומס מיילים גבוה

צריכים סינון חכם, סיקום מהיר של מיילים וכ蒂בה אוטומטית.

## אנשים עם מספר חשבונות מייל

זוקרים לכלי שמרכז ומנהלו את כל החשבונות במקום אחד.

# מתחרים מוביילים

## Outlook

שירות מייל עסקי נפוץ עם التركيز על פרודוקטיביות, ארגון והארגון בו סטטיים ומtabססים בעיקר על תיקיות שנקבעות ידנית.

## Gmail

שירות הדוא"ל הפופולרי ביותר, כולל סינון בסיסי של קידומי מכירות וספאם, אך אינו מציע התאמה אישית או סיווג חכם לפי הקשר תוכן.

از לכל מי שהמשפט

# ניהול פשוט וברור של מיילים

מדובר אליו...

**Falcon**

# פתרונות לניהול חכם של מיילים

*Falcon* היא אפליקציה חכמה לניהול וסינון מיילים בעזרת AI, המבוססת על רשתות נוירוניים שמנתבות ומסוגת מיילים לפי תיוגים מותאמים ולפי רלוונטיות.



סטטיסטיות על השימוש  
במיילים לשיפור  
הפרודוקטיביות



סיכום וכתיבת מיילים בצורה  
תמציתית באמצעות AI



תיוג אוטומטי וסינון לפי  
העדפות אישיות



ריכוז וניהול מסpter  
חשבונות מייל במקומות אחד

Compose with AI

The results! You can edit this, or try again if you like.

Generated Subject:

The subject of the email the AI gen.. |

Generated Content:

Dear Shlomi,  
I wanted to ask you if you can meet tomorrow morning at 09:00.  
Let me know soon.  
Thanks,  
Moshe

**Done**

[Try generate again](#)

account.

10:30

Delete

nsectetur ipsum adipiscing...

10:30

nsectetur ipsum adipiscing elit...

10:30

nsectetur ipsum adipiscing elit...

10:30

nsectetur ipsum adipiscing elit...

Falcon

Take your *Inbox* to new heights

Emails Sent Weekly

5 +0.6%

Emails Received Weekly

10 -0.2%

Spam Emails Weekly

17 0%

Deleted Emails Weekly

3 2.5%

Graphical Overview

Monthly Email Categories

Category	Percentage
Work	10.33%
School	4.19%
Finance	25.33%
Marketing	10.33%
Design	4.19%

Monthly Email Categories

**analytics** **inbox** **compose**

Sent emails

The emails that you sent from your account.

Hila Tsivion

Subject goes here

Norem ipsum dolor sit amet, consectetur

Social School Work

Hila Tsivion 2

Subject goes here

Norem ipsum dolor sit amet, consectetur

School

hilatsivion

Subject goes here

Norem ipsum dolor sit amet, consectetur

Work

Hila Tsivion

Subject goes here

Norem ipsum dolor sit amet, consectetur

Social School Work

analytics inbox compose

Spam

Spam mails will be deleted after 30 days.

Senders Name 11:34

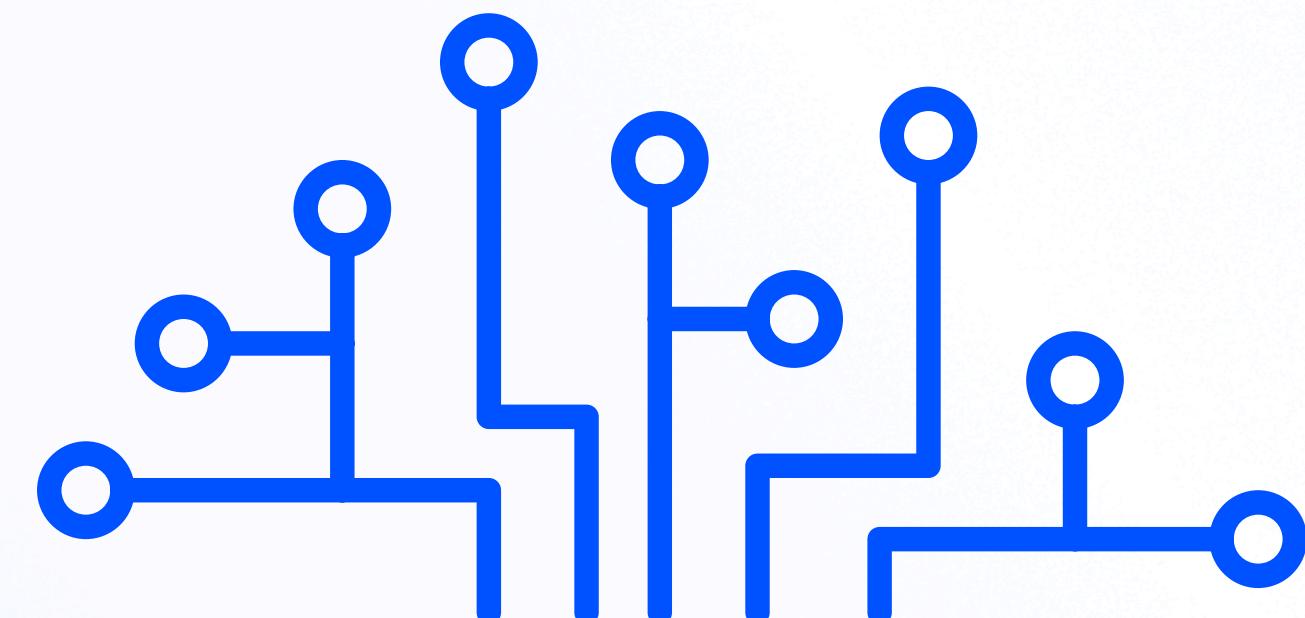
Subject goes here

Norem ipsum dolor sit amet, consectetur ipsum a...

Move to inbox Delete

AI

# רשתות הנוירוניים שלנו



# תיזים אוטומטיים של מיילים

**מייל עבודה? איש? פרסום? לנו להחליט**



יצרנו דата סט של מעל ל-**350,000** טקסטים מיילים שונים המתויגים בתגים רב קטגוריאליים, המאפשרים **לבינה המלאכותית** שלנו ללמידה כיצד זה הות עבר טקסט מסוים מהתיז המתאים לו.

## דיזהוי ספאם ופישינג

**AIR אנהנו מזהים מיילים לא רצויים? באמצעות AI כמובן!**



המיילים שלנו **מוצפים** בפרסומות, הונאות ופניות לא רלוונטיות.

רשת הננוירוניים מאמנת לזהות דפוסים ייחודיים במיילים חסודים,  
תור ניתוח ניסוח תוכן המail.

# הישגי מודל הספאם

## בעזרת Matrix Confusion

טבלה זו מציגה את ביצועי המודל בזיהוי הודעות ספאם ולא ספאם באופן השוואתי.

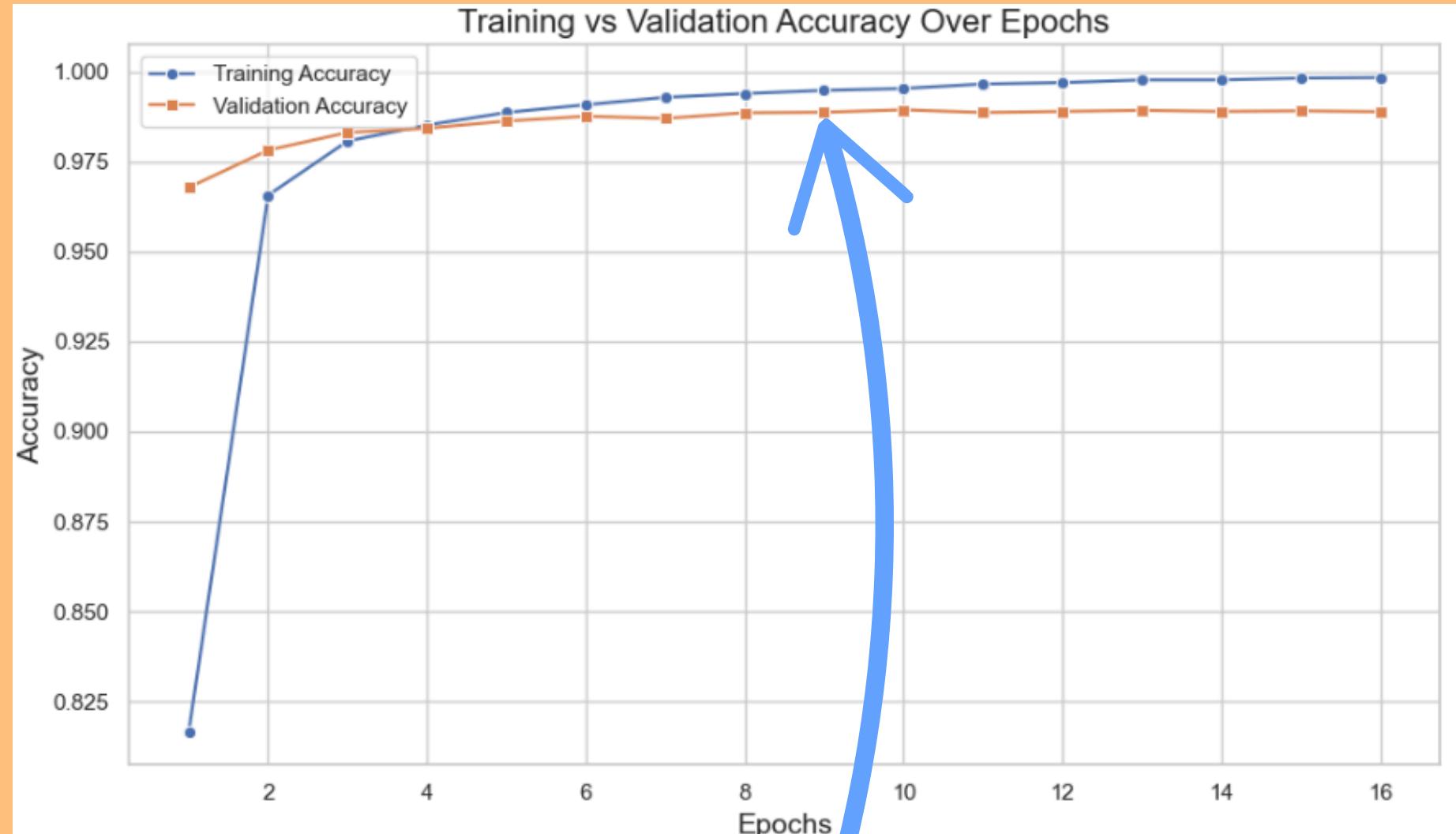
		Predicted Ham	Predicted Spam
Acutal Ham	2,455	31	
Acutal Spam	43	2,307	

( מיילים לגיטימיים = Ham )

**98.6%**

דיק המודל





לכן, בחרנו להשתמש בגרסתה של שלב 9, שבה המודל כלל יותר ופחות תלוי נתונים האימון.

## שיפור דיק מודל הספאם

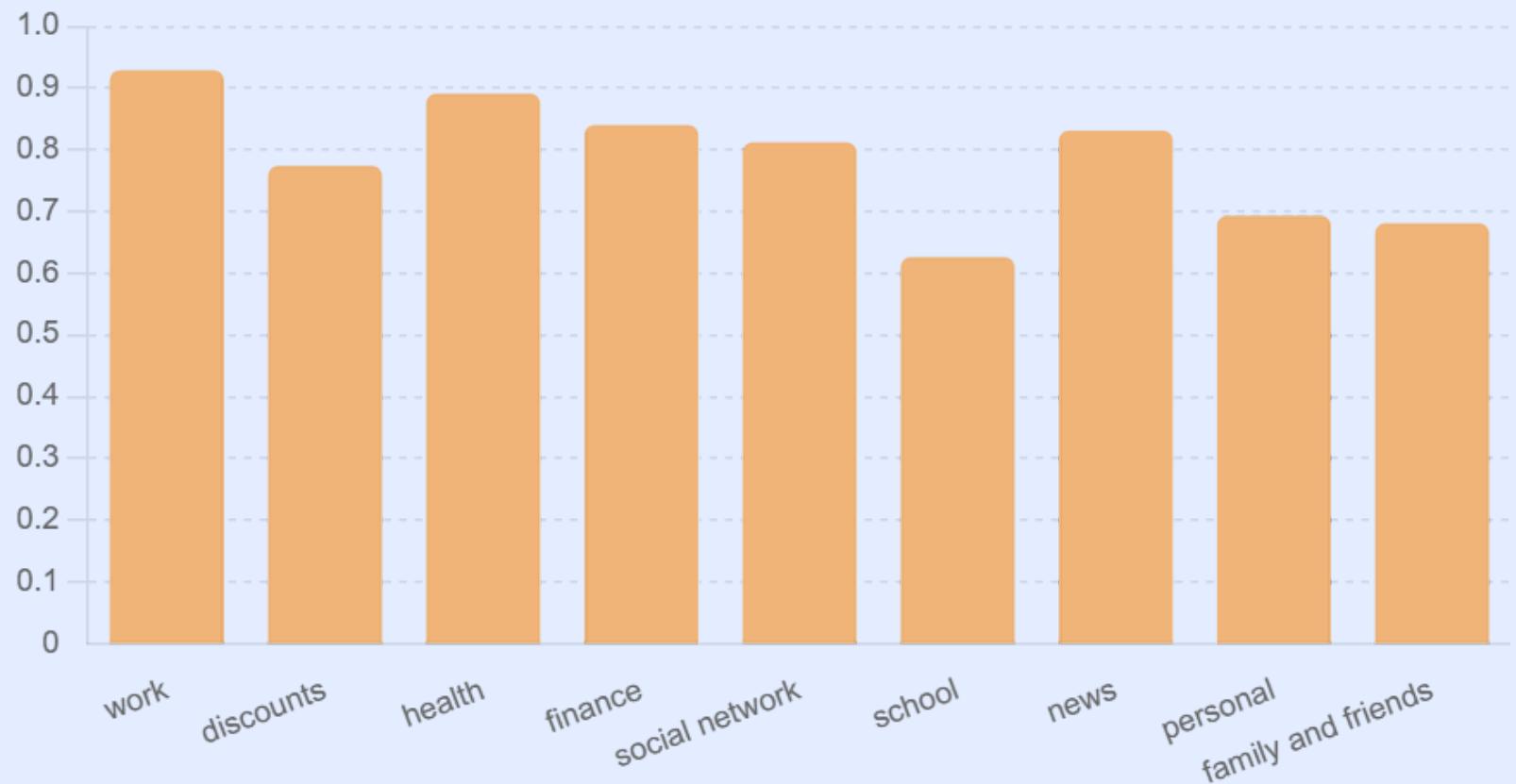
הgraf מציג את שיפור המודל לארך 16 מחרורי אימון.

בהתחלת הדיק (הקו הכהול) עולה במהירות.

סיבב מחזור 9, שיפור הדיק נעצר (הקו הכתום), מה שמעיד על מעבר משלב **למידה לשינוי**.

# הישגי מודל הקטגוריות

## הערכתה אינדיבידואלית



המודל ממיין מיילים לפי נושאים כמו עבודה,  
בריאות וכיספים.  
בגרף מוצגים ציוני F1 בקטגוריות השונות.

**84%**   
דיקן המודל

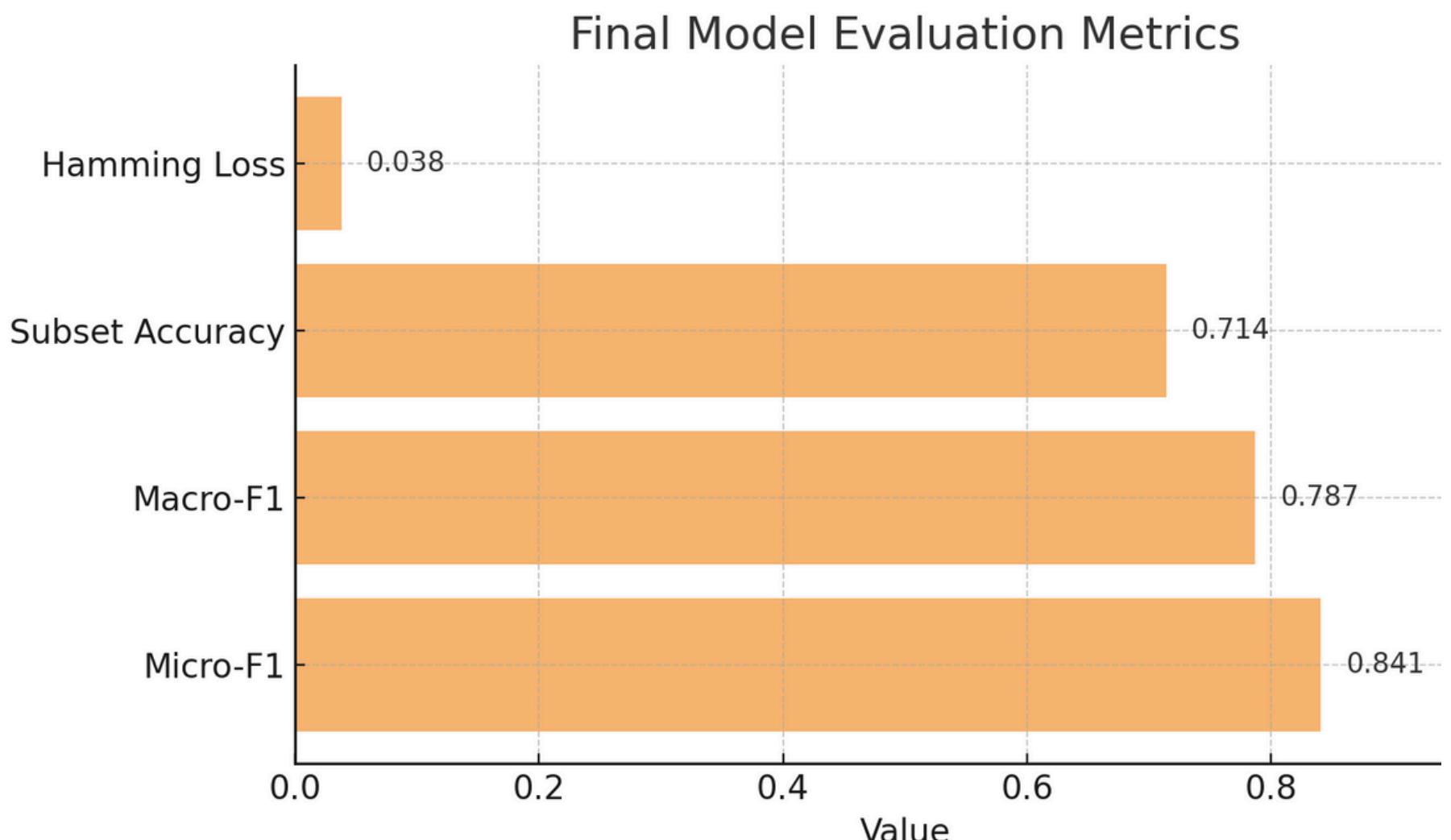
# הישגי מודל הקטגוריות

## הערכת קבוצתית

הגרף מציג את מדדי הביצועים המרכזיים של המודל.

מכיוון שהמודל הוא **multi-label**, כל מייל יכול להשתאיר במספר קטגוריות.

לכן, במקום לבחון רק את מדד ה- $F_1$  לכל קטgorיה בנפרד, חשוב להתמקד במדדים שמעריכים את הביצועים הכלולים של המודל על פני כל הקטגוריות יחד.



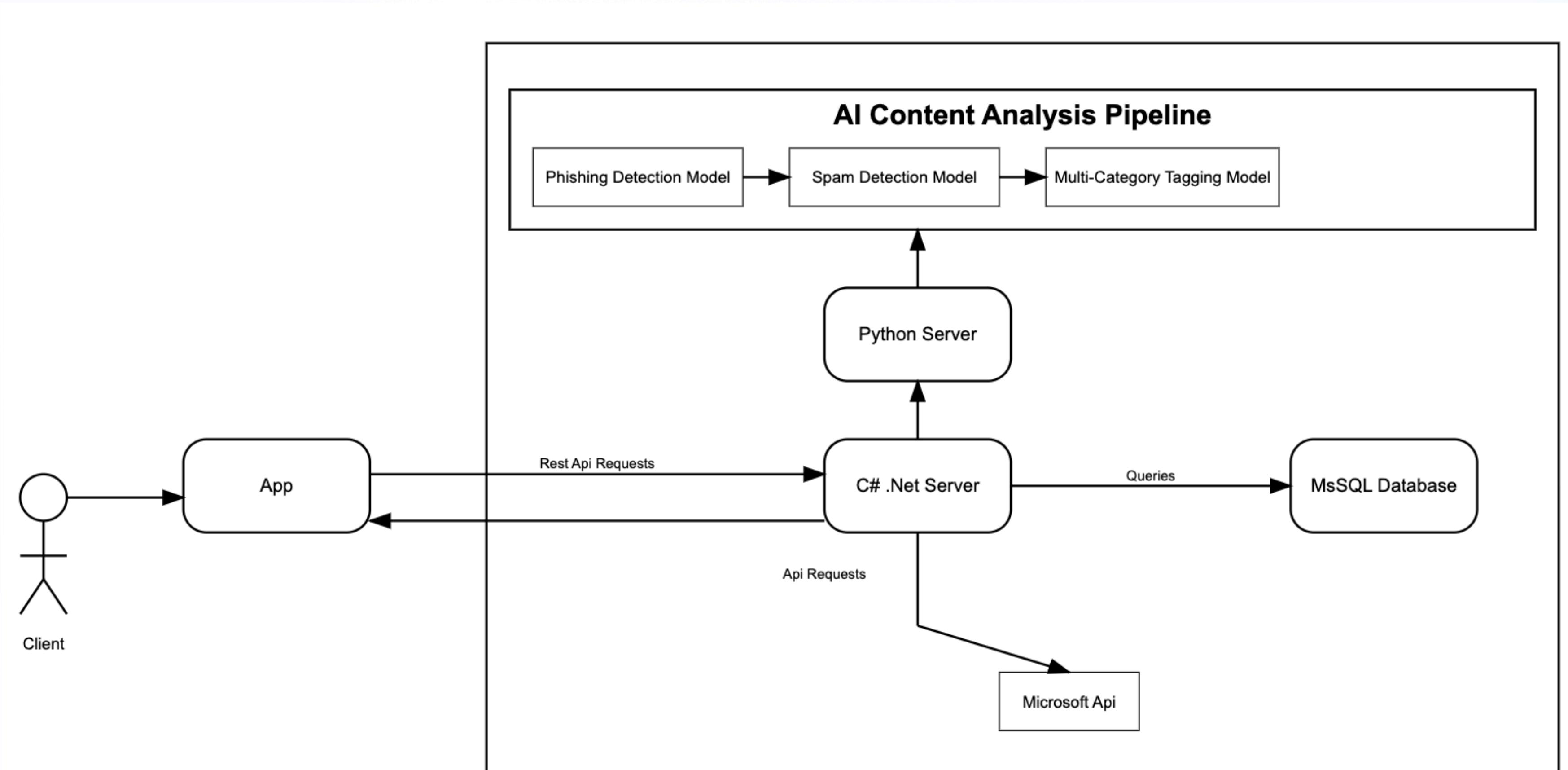
F

צד השרת והדעתה

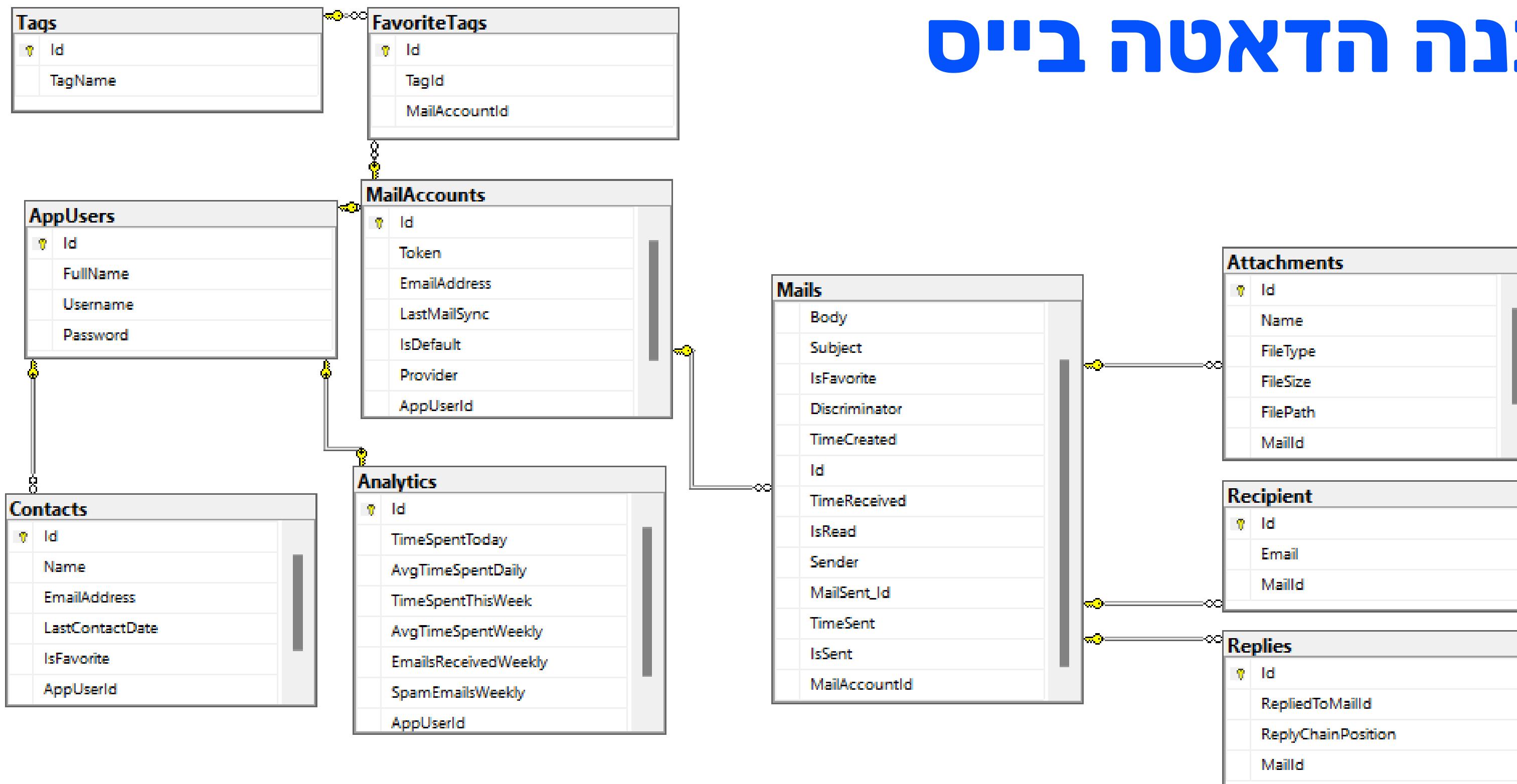
---

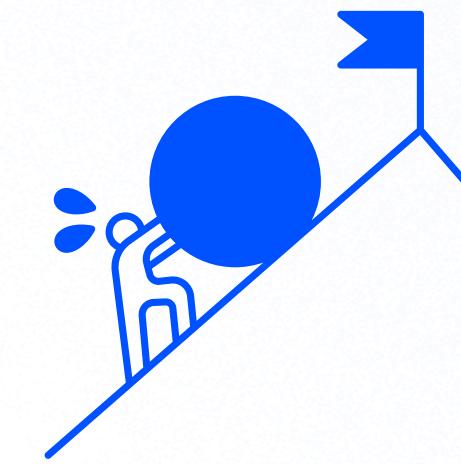


צד השרת והדעתה



# מבנה הדadata ב-YIS





# אתגרים ופתרונות

# התמודדות עם חוסר איזון בנתוניים

## דוגמה לאתגר שחוינו

במהלך פיתוח המודל נתקלנו בכבר של חלק מהקטגוריות היו מעט מיילים, מה **שפגע** ביכולת של המודל ללמידה ולבצע הערכה מדויקת לגבייהן.

בנוסף, קיימ סיכון שקטגוריות נדירות לא ייכללו כלל בקבוצת המבחן, מה שעלול להטעות את התוצאות.



# התמודדות עם חוסר האיזון בנתוניים

פתרונות

כדי לוודא שכל קטגוריה תייצג גם בקבוצות המבחן והולידציה, השתמשנו בשיטת **מחלקת את הדאטה** (iterative stratified splitting) כר שכל קבוצה משקפת את התפלגות הכוללת של הקטגוריות. כר הצלחנו להתמודד עם חוסר האיזון ולבצע הערכה מהימנה.



# מה עבד טוב?

**מה עשינו על מנת להצליח:**

חלוקת תפקידים ואחריות ברורה בין חברי הצוות



עדכוניים שבועיים של סטטוסי התקדמות



מחקר רב ורבה ניסוי וטעייה



**F**

# שאלות מהקהל

תודה על ההקשבה