

פירוט השאלות שהתקבלו, מענה והבהרות לפנייה מוקדמת לקבלת מידע (והדגמה - במידת הצורך) (RFI/RFD) 22/2023
אודות יכולות פיתוח קוד תוכנה ובניית מודלים של למידת מכונה
לצורך סימול אוטומטי ואינטראקטיבי של ענף כלכלי ומשלח יד בסקרי משקי בית ונפשות של הלמ"ס ובמפקד עבור הלמ"ס

להלן פירוט השאלות שהתקבלו ל-RFI, ונוסח מוצע על ידי עורך ה-RFI של תשובות ההבהרה:

#	סעיף	שאלה	תשובת עורך הפנייה
1.		מה הטכניקות שבשימוש בסימול הנוכחי (ML, parsing)?	במערכת הקיימת כיום בסימול, יש מערכת אוטומטית ישנה שמשמשת במילות מפתח ותנאים למתן סמל ענף כלכלי ומשלח יד (שיטת ה " if... then...). הצלחות של מערכת זו ניתן לראות בטבלה בסעיף 4.6.
2.		האם הלמ"ס מעוניינת להשתמש בפתרונות בענף?	ניתן להציג במענה לפנייה זו פתרונות הכוללים שימוש בענף (במקרה זה יש להציג גם פתרונות לסינתוז או להתממה של נתונים באופן שלא ניתן יהיה לזהות את נושאי המידע. הסינתוז או ההתממה יבוצעו ברשת מבודדת עוד טרם השימוש בענף). כמו כן, ניתן להציג במענה לפנייה זו פתרונות מקומיים (On Premise) ברשת מבודדת.
3.		האם הלמ"ס מעוניינת להקים יכולת ML מקומית (חומרה, כ"א בתחום ML, אנליסטים)?	כן
4.		באיזה תכיפות מופיעות דרישות חדשות לתינוג?	כל הזמן קיים חומר חדש שמגיע מהשטח בסקרים שונים או בקבצים מנהליים. לכן תדירות עדכון מודלים ML לפיענוח, תיקבע בזמן הפיתוח והיישום.
5.		האם הלמ"ס מעוניינת לקבל הנ"ל כשירות חיצוני? אם כן, באיזה רובד (משאבי מחשב, תחזוקת מודל, תמיכה בהפעלה שוטפת)?	בשלב RFI הלמ"ס מעוניינת לבחון מגוון אפשרויות למימוש
6.		האם יש בלמ"ס כיום עם ידע ב-ML?	כיום בלמ"ס קיים צוות מפתחי PYTHON . פיתוח יכולות בתחום ML נמצא בשלב התחלתי .
7.		מהם נפחי ההרצה הצפויים (קצב קליטת מידע חדש)?	הקצב שונה בהתאם לתדירות ביצוע הסקרים או קבלת נתונים מנהליים. הקצב יכול להיות אלפי מקרים כל לילה ומצד שני עשרות אלפים פעם בחודש או מאות אלפים פעם בשנה.
8.		האם הסיווג צריך להתבצע ב-BATCHES או ב-REAL TIME בזמן דגימה, ראיון ?	הלמ"ס מעוניינת לראות פתרונות, אם קיימים כאלה, גם לסיווג אחרי קליטת חומרים דרך ממשק סימול במשרד וגם פתרונות, אם קיימים, לסיווג בזמן אמת תוך כדי מילוי שדות בשאלון בזמן הראיון.

~ 1 ~

#	סעיף	שאלה	תשובת עורך הפנייה
9.		האם קיימים/ יהיו אנליסטים בלמ"ס שיעסקו בשיפור המודל?	יוחלט בהמשך האם הלמ"ס תעסיק עובדים לטובת הנושא או/ו תשתמש בשירות חיצוני
10.		מה ההיקף שניתן להשקיע ע"י אנליסטים באופן שוטף, לשיפור איכות המודל ?	הלמ"ס מבקשת כי יוצעו לה פתרונות אפשריים, לרבות היקפי כ"א הנדרש. אופן העסקה – אם באמצעות העסקה ישירה או שאינה ישירה תיבחן בהמשך.
11.	4.2	מבקשים לקבל דוגמאות של מסמכי הבסיס אותם המערכת צריכה לפענח. על הדוגמאות לייצג את התרחישים האופייניים.	לא מדובר על מסמכים. המערכת מקבלת נתונים מסקרים שונים דרך ממשק. השדות שמקבלת המערכת נמצאים בקובץ דוגמאות שצורף לפרסום. העבודה היא מול בסיס נתונים.
12.	4.2	מהו הקלט המתבקש לעיבוד? מסמכים, מאגר נתונים, קבצי טבלה או אחר. נבקש לקבל דוגמאות.	הקלט או תוצאה של העיבוד היא סמל מפורט ברמה של 4 ספרות של ענף כלכלי לפי הסיווג האחד של ענפי כלכלה וסמל מפורט ברמה של 4 ספרות של משלח יד לפי הסיווג האחד של משלחי יד. בנוסף, אמורים להיות גם מספר משתנים טכניים כגון: תאריך סימול, סטאטוס רשומה וכדומה.
13.	4.8	"סימול אוטומטי ברמה גבוהה, מעל 90% - " ניתן לכמת את התוצאות רק ע"ב מידע מופענת. נשמח לקבל dataset מוסכם לצורך בחינת ביצועי המערכת והדגמתה (לביצוע benchmarking).	לא ניתן להעביר dataset מחוץ ללמ"ס.
14.	4.8	"סימול אוטומטי ברמה גבוהה, מעל 90%" - על פי מה נמדד שיעור ההצלחה: 90% מתוך השורות (כאשר שורה המופענת ב-100% נכללת כהצלחה) או על פי שדות (90% מתוך כלל השדות בפיזור כללי).	ההצלחה של פענוח לפי מודל ML נמדדת באמצעות מדדים של טיב ההתאמה מול קובץ תרגול. המדד הוא מתוך מודלים של ML Accuracy שצריך להיות מעל 90%. דוגמה של חישוב הצלחה: יש 1000 רשומות, המודל אמור לספק סמל ענף כלכלי וסמל משלח יד בו זמנית ברשומה כאשר הסמל נכון בביטחון של 90% גם בענף כלכלי וגם במשלח יד בלפחות 500 רשומות מתוך 1000. כמובן שכלל שיהיו יותר מ-500 רשומות מופענות בהצלחה כך זה יהיה יותר טוב למערכת.


#	סעיף	שאלה	תשובות עורך הפנייה
15.	4.11 (4)	נודה לקבלת משמעות ודוגמה להגדרת "מישחוק ותחרותיות".	הכוונה היא להפוך את עבודת הסמלנים ליותר נוחה ופחות מתישה. כיום הסמלנים צריכים לשבת שעות רבות מול המסך ולפענח רשומות.
16.	4.12	האם שאלוני הלמ"ס הם דיגיטליים?	מבחינת מערכת הסימול, כל החומר מגיע דרך ממשק באמצעות טבלאות SQL. כלומר, בפורמט דיגיטלי. זה לא משנה איך נאסף החומר בסקר ספציפי.
17.	4.12	נודה לקבלת דוגמאות לפלט לשאלונים הרלוונטיים.	הדוגמה קיימת בנספח של המשתנים שמגיעים לממשק סימול לצורך פיענוח.
18.	4.12	האם הקלט לעיבוד (התשובות לשאלונים) מכיל כתב-יד? במידה וכן, נשמח לקבל דוגמאות לאמור.	לא רלוונטי. החומר לסימול מגיע דרך ממשק מתוך טבלאות SQL.
19.	4.12	האם תיאור הסמלים עליו מתבסס הסיווג נמצא בקישורים המצורפים לסעיף 4.3?	כן
20.		בשל עבודה עם צוותים גלובליים ומידע רב שקיים בשפה האנגלית, נבקש להגיש את המענה בשפה האנגלית.	ניתן לבצע הצגה באנגלית. אבל הלמ"ס תצטרך לקבל הדגמות של יכולת עיבוד שפה עברית – חובה, כי כל החומרים בלמ"ס הם בעברית בלבד.
21.	4.13 (1)	"על הספק להציע את פתרונות לרכיבים 3, 4, 5, ו-7 של יצירת תהליך": האם אפשר לתאר עם יותר פירוט מהן הציפיות והדרישות עבור רכיבים 5 ו-7?	ראה תשובה לשאלה 14. רכיב 7 – הלמ"ס מצפה למערכת ניהול סימול שתכלול גם דו"חות מעקב על ביצוע תהליך הסימול וגם מסמכים לעבודה ידנית (סימול ידני) ותמוך מחשב של סמלנים. לאחר סיום תהליך ה-RFI, בהתאם להחלטת הלמ"ס להמשיך התהליך, הלמ"ס תשקול להגדיר דרישות מדויקות יותר, אשר יוגדרו בהמשך במסמכי דרישות, במידה ויפורסמו כאלה. בשלב זה, הלמ"ס מבקשת ללמוד מיכולת המציעים לפתח מערכות ניהול של תהליכים.
22.	4.13 (2)	"העבודה בחצרות למ"ס תהיה בסביבה מנותקת מרשת האינטרנט", מה הכוונה? האם סביבת ענן מאובטחת על IP-private שאינה חשופה לרשת האינטרנט מקיים את הדרישה?	ראו תשובה לשאלה 2.

תשובת עורך הפנייה	שאלה	סעיף	#
זה תלוי בפיתוח המודלים. אם בונים מודלים נפרדים לענף כלכלי ולמשלח יד אז מדובר על 4 מודלים. אם בונים מודלים שמפענחים גם ענף כלכלי וגם משלח יד בו זמנית, אז מדובר בשני מודלים. העיקר הוא שיהיו לפחות 2 מודלים שעובדים בשיטות שונות ובלתי תלויות אחת בשנייה, כדי לאמת את הסמלים המתקבלים גם של ענף כלכלי וגם של משלח יד.	האם מדובר בסך הכל על 4 מודלים שונים?	4.13 (3)	23.
מערכות בתצורות שונות מאפשרות ביצועים התלויים במידול ובארכיטקטורה. מהירות התגובה של המערכת תלויה בכמות הרשומות/ מודל/ שאיגום מידעיים מסוגל לפענח, למפות ולהציג (כחלק מטרנספורמציה המידע) עבור יחידת זמן. המערכת נדרשת להיות בעלת יכולת של אופטימליות בטכניקות הפילוח של המידע, על מנת לאפשר ביצועים מקסימליים. טכניקות הפילוח ואיגום המידע תאפשרנה איגום אופטימלי של מערכים וקבוצות מידע.	מה הכוונה של : "עמידה בעומסים בתצורות שונות"? "טכניקות פילוח"?	4.14 (1)	24.
בשלב זה, תצורת עבודה מקומית On Premise - לא בענן	המשיב מתבקש לתאר את הארכיטקטורה ליישום המערכת בתצורה מקומית. מה זאת אומרת תצורה מקומית?	5.2 (1)	25.
בשלב זה, תצורת עבודה מקומית On Premise - לא בענן. ראו גם תשובות לשאלות 2 ו-22	כיצד מתבצע ממשק לשירותי outlook? איזה ממשקים נדרשים לשירותי Outlook? האם יותקן Outlook בענן? נשמח אם תוכלו להרחיב על רכיב ה-outlook, מה תפקידו בארכיטקטורה והאם ניתן להחליפו במוצר אחר?	5.4 (2)	26.

הערות:

בעקבות השאלות. שהתקבלו ל-RFI, אין שינוי במסמכי הפנייה המוקדמת לקבלת מידע (RFI).

חתימת עורך הפניה

שם	אריאל קדושי
תאריך	3/7/2023
חתימה	

200ak23