

שימוש בבינה מלאכותית לשיפור תפקודו של מערך הדיפלומטיה הציבורית בישראל

אופיר בראל

אוניברסיטת תל אביב

הופעת ה-ChatGPT עוררה ברחבי העולם דיונים ערניים בדבר היתרונות והחסרונות המשמעותיים הטמונים בו. מערך הדיפלומטיה הציבורית בישראל יכול להטמיע את השימוש בבינה מלאכותית, לרבות ChatGPT, על מנת לשפר את הליכי התכנון והביצוע של פעולות לשיפור תדמיתה של ישראל בעולם. השיפור עשוי להתבטא בשלוש רמות: ברמה האסטרטגית – הגדרה מחדש של בריתות בינלאומיות לאור שינויים דמוגרפיים עולמיים, הטומנים בחובם השלכות פוליטיות ומדיניות; ברמת הביניים – שיפור הליכי התכנון והתיאום הבירוקרטיים בין הגופים הרבים המרכיבים את מערך ההסברה, על ידי שיפור תהליכי הקצאת המשימות ויצירת שקיפות; ברמה הטקטית – שיפור המענה למבצעי השפעה אנטי-ישראליים ברשתות החברתיות על ידי הפקה מהירה של תגובות ממוקדות.

מילות מפתח: ChatGPT, דיפלומטיה ציבורית, בינה מלאכותית, דיסאינפורמציה, ישראל

מבוא

הצגת הצ'טבוט ChatGPT על ידי חברת OpenAI בנובמבר 2022 עוררה הדים רבים בעולם בשל יכולותיו המתקדמות, שעשויות לשפר את איכות החיים של בני אדם ([למשל](#) לדיוק האבחון הרפואי) או להיות מנוצלות למטרות זדוניות ([למשל](#) לביצוע מתקפות דיג יעילות יותר בקלות רבה). התפתחות זו היא בעלת משמעויות פוטנציאליות ניכרות גם בתחום היחסים הבינלאומיים ובייחוד בתחום הדיפלומטיה הציבורית, מאחר שהיא מדגישה את מקומו ההולך וגדל של מרחב האינטרנט בניהול יחסים בין מדינות, עמים ואנשים – מגמה המתרחשת מזה כשני עשורים. שימוש בבינה מלאכותית להפקת סרטונים מסוג זיוף עמוק (deepfake), [כפי שנעשה למשל](#) במסגרת המלחמה באוקראינה, עלול להשפיע על תפיסת הציבור בנוגע לאירועים מסוימים במלחמה ואף על מהלכה הכולל.

ישראל ניצבת מיום הקמתה בפני אתגר מתמשך באשר לניהול תדמיתה בעיני העולם, ועל כן השקיעה במהלך השנים מאמצים ומשאבים רבים בתחום ההסברה. למרות זאת, רבים בציבור [תופסים](#) את תפקודה של ישראל בזירת התודעה הבינלאומית כלקוי, בשל קיומן של תנועות אנטי-ישראליות רחבות בעולם המנסות לפגוע בלגיטימיות שלה כמדינה עצמאית.

ההתפתחויות הטכנולוגיות בתחום מחד גיסא, והעניין הציבורי בישראל בתפקוד הדיפלומטיה הציבורית מאידך גיסא, יוצרים עבור מערך ההסברה הזדמנות לקדם הטמעה של שימוש באמצעי בינה מלאכותית לשיפור ביצועיו.

השימוש בבינה מלאכותית בדיפלומטיה הציבורית טומן בחובו הזדמנויות וסיכונים רבים שיפורטו בהמשך. ההתפתחויות הטכנולוגיות בתחום מחד גיסא, והעניין הציבורי בישראל בתפקוד הדיפלומטיה הציבורית מאידך גיסא, יוצרים עבור מערך ההסברה הזדמנות לקדם הטמעה של שימוש באמצעי בינה מלאכותית לשיפור ביצועיו. השיפור עשוי להתבטא בשלוש רמות: הרמה האסטרטגית – השימוש בבינה מלאכותית יאפשר לישראל לחזות באופן טוב יותר שינויים גלובליים עתידיים, שעשויים לשנות את מאזן העוצמה העולמי ולהשפיע על האופן שבו תעדיף ישראל לנהל את יחסי החוץ שלה, בייחוד עם המעצמות; רמת הביניים – הבינה המלאכותית עשויה לתת מענה לסבך הבירוקרטי הקיים במערך ההסברה, שמונע לא פעם ממבצעי השפעה ישראליים לפעול באופן יעיל; הרמה הטקטית – הבינה המלאכותית תאפשר לגורמי ההסברה בישראל לתת מענה מהיר

הדבר ניכר בראש ובראשונה בקהלי היעד בארצות הברית. הדיפלומטיה הציבורית הישראלית **מעדיפה** לפנות לאלה המזוהים באופן מסורתי כפרוידישלאים (אוונגליסטים ויהודים) על פני צעירים אפרו-אמריקאים והיספנים אמריקאים. למעשה תמונת המצב מורכבת יותר. מחקרים מהשנים האחרונות מראים כי לצד שיעורים גבוהים של אזרחים בקבוצות הללו, **האוחזים בתפיסות אנטישמיות**, יש גם **תמיכה ואהדה** בלתי מבוטלות כלפי ישראל והיהדות. במישור הדמוגרפי יש לציין כי בעוד שיעורם של האונגליסטים באוכלוסיית ארצות הברית **ירד** בין השנים 2007 ל-2021, הקהילה ההיספנית היא האוכלוסייה **הצומחת בקצב המהיר ביותר** במדינה. למרות זאת, במקרים שבהם **מיושמת** פעילות כלפי הקהילה ההיספנית היא נקודתית, ורובה אינה מבוצעת בשל מחסור תקציבי.

ברמה האסטרטגית ניתן להשתמש ב-ChatGPT להפקת תחזיות דמוגרפיות מגוונות ולניתוח מעמיק של משמעותיותן המדיניות. בחקר שינויי האקלים, לדוגמה, ChatGPT **מזהה** כאמצעי המסוגל לקדם את הבנת טיבם של השינויים ולשפר את הדיוק של תחזיות האקלים, תוך שימוש בנתונים המגיעים ממקורות רבים. לצד זאת ניתן להשתמש ב-ChatGPT גם ליצירת תרחישים בעלי רמת דיוק גבוהה, המתארים את הנגזרות הפוליטיות והמדיניות של השינויים הדמוגרפיים החזויים. יצירת תרחישים כאלו מחייבת את המשתתפים להפגין יכולות תכנון וביצוע גבוהות, והיא מוכרת כבר כיום כמרכיב אפשרי **בהכשרת צוערים**.

ניתן גם להיעזר ב-ChatGPT לצורך היכרות מעמיקה עם האידיאולוגיות והערכים המכוננים של קהילות היעד הנבחרות. בארצות הברית למשל, תפיסותיהן של האוכלוסיות הלא-לבנות **שונות** מאלו של האוכלוסייה הלבנה, שעמה התרגלה ישראל לעבוד במהלך השנים. אי-הכרת עם ההבדלים עלולה להוביל ליישום **תפיסות שגויות** ביחס לקהילות אלו בעת תכנון וביצוע ההסברה הישראלית, ובכך עלול תפקודה להיפגע.

ניתן להכיר לעומק את קהילות היעד בעיקר על ידי שימוש ב-ChatGPT לצורך ניתוח סנטימנט (Sentiment Analysis) וניתוח עמדות (Stance detection). ניתוח סנטימנט הוא שימוש בטכניקות עיבוד שפה מבוססות בינה מלאכותית לכריית טקסטים ולניתוחם, לשם זיהוי וכימות מצבים רגשיים שכותביהם מביעים. **ניתוח עמדות** נחשב ליכולת מתקדמת יותר, מאחר שהוא מאפשר לא רק לאפיין את הרגשות המובעים בטקסט אלא גם לאפיין את עמדת הכותבים ביחס לנושא מסוים. בשל כך ניתוח עמדות **נחשב** לכלי מרכזי בניתוח דעת הקהל ברשתות החברתיות – בייחוד בסוגיות פוליטיות

ואיכותי יותר לפעולות דיסאינפורמציה ודה-לגיטימציה המתנהלות נגדה ברשתות החברתיות – לרבות פעולות המתבססות בעצמן על שימוש בבניה מלאכותית. בסוף המאמר מוצגות מספר המלצות לפעולה, שעשויות להתחיל את תהליך ההטמעה.

הרמה האסטרטגית: הבנה מעמיקה של שינויים דמוגרפיים ומשמעותם המדינית

הדיפלומטיה המסורתית מתבססת אומנם בעיקר על פעילותם של נציגי מדינות רשמיים, אולם האינטרנט והרשתות החברתיות הפכו גורמים לא-ממשלתיים, חברות עסקיות וקבוצות חברתיות לבעלי השפעה רבה יותר בעיצוב היחסים הבינלאומיים. בשל כך שינויים חברתיים ודמוגרפיים המתחוללים באוכלוסיות מסוימות נעשים משמעותיים יותר בהשפעה על מטרות הדיפלומטיה הציבורית, לצד השיקולים הקלאסיים של מאזני עוצמה בין מדינות והאינטרסים הלאומיים שלהן. שינויים דמוגרפיים עשויים להתבטא בירידה או בעלייה במספר תושבי המדינה (**לדוגמה**, אוכלוסיותיהן של סין ואירופה – הנחשבות כיום למעצמות כלכליות – צפויות להצטמצם עד שנת 2100), או בשינוי תמהיל האוכלוסייה המקומית (אוכלוסיות מהגרים תופסות נפח הולך וגדל באוכלוסיותיהן של **אירופה וארה"ב**). צמצום או גידול בנפח האוכלוסייה עשוי להשפיע באופן ישיר על מרכיבי עוצמה קשה של מדינות (למשל על ידי גידול או פיצוח בהיקף האוכלוסייה בגיל העבודה), בעוד ששינוי הרכב האוכלוסייה עשוי לגרום לשינויי מדיניות בשל שינויים בדפוסי ההצבעה, שעשויים להוביל לעליית ממשל המציג גישה שונה במדיניות חוץ.

שינויים דמוגרפיים מקבלים משמעות מיוחדת בניהול פעולות הדיפלומטיה הציבורית של מדינות קטנות, בהן ישראל: אותן מדינות מיישמות **שני אמצעים** מרכזיים לשמירה על עוצמתן בזירה הבינלאומית: האחד – היכולת לשכנע מעצמות גדולות לקדם עבורן פעולות המממשות את האינטרסים הלאומיים שלהן; השני – יצירת קואליציה של מספר מדינות קטנות המקדמות אינטרסים משותפים. גורמים רשמיים ולא-רשמיים המעוניינים לפעול למימוש מטרות אלו **נדרשים**, מצד אחד, לייסד ולממש קשרים קרובים עם מקבלי ההחלטות הרשמיים, ומצד שני לרתום לפעולה קהילות אזרחים פיזיות ומקוונות.

מטבע הדברים, שינויים דמוגרפיים מחייבים את אותם גורמים להכיר טוב יותר את אופי השינויים ואת השפעותיהם. גורמים דיפלומטיים בישראל אומנם זיהו **את הצורך להיערך לשינויים דמוגרפיים** עולמיים, אך לא נראה כי הם אכן מביאים לשינוי בהגדרת קהלי היעד.

המורכב מגופי ממשלה רבים, שקיבלו ברבות השנים סמכויות רבות **על חשבון משרד החוץ**. אומנם מטה ההסברה הלאומי **מונה** להיות הגוף האחראי על תיאום פעולות ההסברה של ישראל, אך בפועל הוא **מתקשה** בביצוע משימתו, **ולמעשה** פעילות ההסברה נעשית ללא בסיס ניהולי מרכזי המגדיר את מטרותיו האסטרטגיות של המערך. בעיות תכנון קיימות גם בעבודת משרדי הממשלה עם גופים חיצוניים. דוגמה לכך היא מיזם בשם "מורשה בניין" לחיזוק ההסברה הפרו-ישראלית במוסדות אקדמיים בארצות הברית, **שנוהל על ידי משרד החוץ ועמותה חיצונית** מבלי שנבחנה השגת יעדי, וחרף הערכה פנימית כי התועלת שהושגה ממנו אינה מצדיקה את העלות הגבוהה שהושקעה בו.

השימוש ב-ChatGPT עשוי לסייע לישראל לשפר את התכנון, היישום והעדכון בזמן אמת של משימות ההשפעה הנגזרות מהתכנון האסטרטגי.

השימוש ב-ChatGPT עשוי לסייע לישראל לשפר את התכנון, היישום והעדכון בזמן אמת של משימות ההשפעה הנגזרות מהתכנון האסטרטגי. ניתן לעשות זאת **למשל** באמצעות מתן המלצות לחלוקת משימות מיטבית בין צוות עובדי הפרויקט והצגת הערכות בדבר שינויים שיש להתחשב בהם (למשל עיכובים אפשריים בהשלמת הפרויקט). כמו כן ניתן להשתמש ב-ChatGPT **לאיסוף ולסיכום מידע** ממקורות רבים, בייחוד בנושאים הקשורים לשיפור ניהול פרויקטים, כגון מתודולוגיות שונות לניהול פרויקטים. לבסוף, ניתן להשתמש ב-ChatGPT להקמת פלטפורמת תקשורת מרכזית בין כלל הגורמים השותפים בדיפלומטיה הציבורית. פלטפורמה כזו לא רק תצמצם מקרים של אי-הבנות, עקב הגעתו של מידע חלקי לאחד המשתתפים, אלא גם **תיצור שקיפות** של תהליכי קבלת ההחלטות, ובכך תחזק את אמון הציבור והמשתתפים בתפקוד הדיפלומטיה הציבורית.

הרמה הטקטית: שיפור ההתמודדות עם מבצעי השפעה אנטי-ישראליים

לצד התועלות האפשריות ברמה האסטרטגית וברמת הביניים, יש לשים לב גם להשלכות השימוש בבינה מלאכותית על מבצעי השפעה ממוקדים: היא אומנם עשויה להחריף את האיום של הפצת מידע כוזב באינטרנט, אך גם לספק פתרונות משופרים לשם התמודדות עם התופעה.

השימוש בבינה מלאכותית עלול לשפר את היצירה וההפצה של תכנים כוזבים ברשתות החברתיות בשני

וחברתיות. במהלך 2023 פורסמו מחקרים שהשוו את ביצועי ה-ChatGPT במשימות אלו למודלי שפה אחרים, ומהם עולה כי ה-ChatGPT **מסוגל** לבצע ניתוח סנטימנט באופן מדויק יותר מאשר מודלי שפה אחרים. נוסף על כך, בנוגע לניתוח עמדות ה-ChatGPT **נחשב** למודל הטוב ביותר כיום, והוא עשוי אף לשנות את הפרדיגמה המחקרית בתחום.

ברמה האסטרטגית ניתן לטעון אפוא כי השימוש בבינה מלאכותית עשוי לאפשר לכל מדינה לחזות תמורות דמוגרפיות, להעריך ביתר דיוק את השלכותיהן המדיניות ולגבש בצורה טובה יותר את עקרונות הפעולה ההסברתיים והמדיניים הרצויים באותן מדינות. התועלות הללו חיוניות במיוחד עבור מדינות קטנות כישראל: בשל תלותן במעצמות גדולות לשם מימוש האינטרסים שלהן הן מושפעות במידה רבה משינויים במאזן העוצמה העולמי, שתמורות דמוגרפיות עשויות למלא בו תפקיד מרכזי. בד בבד, הבנת משמעותם המדינית של השינויים עשויה להשפיע על החלטתן של מדינות קטנות לקדם בריתות בינלאומיות חדשות עם מדינות אחרות, או לשנות את אופיין של קואליציות בינלאומיות שהן חברות בהן. **באירופה למשל**, מעניין לראות כי אוכלוסייתן של מדינות כגון לטביה, ליטא, הונגריה ופולין – שהיו בשנים האחרונות **יעד מדיני מועדף** על ישראל – עלולה להצטמצם, בעוד האוכלוסייה במדינות כמו שוודיה ואירלנד, שעדיין נרשמו בעשורים האחרונים לא מעט **משברים דיפלומטיים**, צפויה דווקא להמשיך לגדול. הדבר עשוי לעודד חיזוק קשרים עם מדינות בעלות צמיחה דמוגרפית, על חשבון או לצד קידום הברית המזרח-אירופית.

רמת הביניים: שיפור הליכי התכנון והיישום במערך הדיפלומטיה הציבורית

הבנה מיטבית של תמורות אסטרטגיות בינלאומיות וגיבוש מדיניות עבור דיפלומטיה ציבורית תומכת – טובות ככל שיהיו – לא יוכלו לתרום למדינה בזירה העולמית אם הגורמים האחראים ייתקלו בקשיים בירוקרטיים בפן היישומי.

המאמצים הרבים שמשקיעה ישראל בתחום הדיפלומטיה הציבורית כוללים גם שימוש בטכנולוגיות חדשות: משרד החוץ **מקדם** שימוש נרחב ביכולות דיגיטליות מגוונות בדיפלומטיה הציבורית, למשל באמצעות קיום הכשרה אינטנסיבית בתחום הדיפלומטיה הדיגיטלית במסגרת קורס הצוערים – מה שהפך אותו **לאחד ממשרדי החוץ המשפיעים במרחב הדיגיטלי**.

למרות זאת, יכולתה של ישראל להשיג הצלחות בתחום זה נפגעת בשל אתגרים בירוקרטיים. מערך הדיפלומטיה הציבורית בישראל הוא מערך מסועף

משמעות כלל ההתפתחויות הללו היא שמגוון רחב יותר של גורמים עוינים מסוגלים להפיץ מידע כוזב ברשתות החברתיות – בייחוד כזה הנוגע למסריה המרכזיים של ההסברה הישראלית – בכמות ובאיכות גבוהות יותר ועם סיכויי גילוי נמוכים יותר מצד הרשתות החברתיות, גופי אכיפה וגולשים.

ההתמודדות עם הפצת תכנים כוזבים מבוססי ChatGPT היא חלק מפעילות עולמית ענפה בשנים האחרונות, הממוקדת בהתמודדות עם מבצעי השפעה מבוססי בינה מלאכותית. מתוך מגוון היוזמות והאמצעים בנושא ניתן לזהות שלושה אפיקי פעילות אפשריים, שבאמצעותם יכול מערך ההסברה לרתום את ChatGPT לצורך מתן מענה משופר למבצעי השפעה כוזבים, לרבות אלו המבוססים על בינה מלאכותית.

אפיק הפעילות הראשון הוא פיתוח אמצעים טכנולוגיים לזיהוי תכנים מבוססי ChatGPT, ביניהם ניתן לציין את **GPTZero**: אמצעי מבוסס בינה מלאכותית העושה שימוש במודל הפעולה של ChatGPT כדי לזהות אם טקסט מסוים נכתב על ידי אדם או על ידי בינה מלאכותית. **GPTZero** **אומנם** משיג שיעורי הצלחה גבוהים באפיון טקסטים ארוכים, אך מתקשה בניתוח טקסטים קצרים כגון ציוצים ב־X (טוויטר).

אפיק הפעילות השני הוא שימוש ב־ChatGPT **ליצירה והפצה מהירות** של תכנים הסברתיים: פוסטים ברשתות החברתיות, הודעות לעיתונות ואף נאומים רשמיים. משרד החוץ השתמש לראשונה בבינה מלאכותית לצורכי הסברה בחודש ינואר האחרון, **כשנעזר** בפלטפורמה שפיתחה חברת ההזנק הישראלית D-ID על מנת להפיק סרטונים קצרים בשמונה שפות ובהן פרסית, סינית וערבית.

במקרים מסוימים איכות המסרים שהפיק ה־ChatGPT **דומה מאוד** לזו של מסרים המיוצרים על ידי שדלנים (לוביסטים) אנושיים. בפועל קיים חשש שתהליך הפקת התכנים באמצעות ChatGPT יתבצע ללא הקפדה על ניסוחם של התוצרים המופקים. בעולם הדיפלומטיה, שבו לכל ניואנס יש משמעות מדינית, **בחירה בניסוחים לא־מדויקים** עלולה להוביל לפערים בציפיות ובתקשורת בין מדינות, ואף למשבר דיפלומטי ביניהן. לצד זאת, א־הקפדה על ניסוחים **עלולה להוביל לכך** שתכנים מסוימים ייתפסו על ידי קהלי היעד הזרים כקלישאתיים מדי או לא־אותנטיים דיים. ביחס לישראל, השימוש ב־ChatGPT **עלול** להוביל גורמים ישראלים המשתמשים בו להטמיע מונחים שלכאורה נראים ניטרליים, אך למעשה משרתים היטב גופים אנטי־ישראליים. דוגמה לכך היא אימוץ המילה 'תנועה' (movement) על ידי ארגון ה־BDS, המעניק לו תדמית

אופנים מהותיים. ראשית, הוא עלול להרחיב את היקף המידע הכוזב המתפרסם ברשתות, הן על ידי הגדלת מספר התכנים שניתן לפרסם והן על ידי הרחבת מספר הגורמים העוינים המסוגלים לעשות זאת. בעוד שהפצת מידע כוזב ללא בינה מלאכותית מתבססת בדרך כלל על הפצת תוכן מסוים במספר פלטפורמות במקביל – תהליך **שעלול להיות בו** חזרתיות ויצירת שחיקה ועומס קוגניטיבי על יוצריו – השימוש ב־ChatGPT מאפשר לגורמים זדוניים ליצור **גרסאות רבות** של תוכן מסוים ולהפיץ אותן בפלטפורמות מרובות. במקביל ההתפתחויות ב־ChatGPT מאפשרות לגורמים זדוניים **להרחיב במידה משמעותית** את היקף הנושאים שעבורם ניתן ליצור תוכן כוזב.

זאת ועוד, הרחבת היקף הגופים המעורבים בהפצת מידע כוזב מתאפשרת בשל העובדה שאימוץ בינה מלאכותית מוזיל את עלותם של אמצעים מתקדמים ליצירה ולהפצה של מידע כוזב. בהיעדרם נאמדת **עלות ההוצאה לפועל** של מבצע השפעה רחב היקף בעשרות מיליוני דולרים, והדבר מקשה על גורמים זדוניים בעלי משאבים מוגבלים להוציא לפועל מבצע השפעה נרחב ברשתות החברתיות. שימוש בבינה מלאכותית מפחית למשל את עלות יצירתם של תמונות ו**קטעי קול מזויפים** באיכות גבוהה ואת עלותן של בחינות מסוג **cross-platform testing**, המאפשרות לבחון את תגובות הגולשים לתוכן המופק במסגרת מבצע השפעה. עוד טרם הופעת ה־ChatGPT **זוהתה** יכולתם של גורמים אנטי־ישראליים להתאגד ולהפיץ את מסריהם בהיקפים גדולים, מה שהעניק להם בולטות גבוהה ברשתות. הוזלת עלויותיהם של מבצעי השפעה מתקדמים עלולה אפוא להעמיק את הפער הכמותי לטובת אותם גורמים ולחזק את נוכחותם ברשתות, על חשבון גורמים הפועלים לטובת ישראל.

שנית, השימוש בבינה מלאכותית מאפשר לגורמים עוינים **לשפר** את ניסוחם של תכנים כוזבים, מה שעלול לגרום לקהלי היעד לתפוס את המידע הכוזב **כמשכנע**. בפרט יש לציין כי **התוצרים המופקים** על ידי ChatGPT-4 נחשבים למשכנעים ולמפורטים יותר, בהשוואה לגרסה 3.5. ואולם השיפור באיכות התכנים הכוזבים לא רק מעלה את סיכויי להשפיע על קהל היעד הנבחר אלא גם **מקשה** על איתורו והסרתו. לצד זאת יש לציין כי ניתן להשתמש ב־ChatGPT גם על מנת **לשתול בדעת הקהל הנוכחית מידע שגוי** על אודות אירועים היסטוריים, באמצעות זיוף ועיוות מסמכים היסטוריים התומכים בהם. ליכולת זו חשיבות מיוחדת עבור ישראל, המרבה לעשות שימוש בזיכרון השואה ובזכות העם היהודי על ארצו כמסרים מרכזיים בדיפלומטיה הציבורית שלה.

היכולת להשתמש בבדיקת עובדות נגד טענות מצד גורמים אלו, מבלי לעבור הכשרה כבודק עובדות מצד גוף כלשהו. כך ניתן להרחיב באופן ניכר את היקפו הפוטנציאלי של מערך הדיפלומטיה הציבורית ברשתות החברתיות – **צורך שזוהה כחיוני** לאחר מבצע שומר החומות (מאי 2021).

במשימות השפעה מוגדרות יכולים גורמי הסברה להפיק תועלת מהשימוש ב-ChatGPT בעיקר לשם פעילות מהירה ויעילה סביב אירועים נקודתיים (למשל מבצע צבאי). המהירות **היא מרכיב חיוני** בהפרכת מידע כוזב המופץ ברשתות החברתיות, והניסוח מהיר של תגובות ומסרי הסברה לצד הקצאה מיטבית של המשימות הנלוות למבצע ההשפעה עשויים להבטיח את השגת הפעילות המהירה. נוסף על כך ניתן לגייס במהירות גורמים שאינם כלולים בדרך כלל במערך ההסברה (אזרחים), על ידי שימוש במיזמים הדומים ל-SAVEE. השיפור במהירות בזכות כלל המרכיבים הללו עשוי בסופו של דבר לתת מענה מסוים לבעיית הפער הכמותי לטובת גורמים אנטי-ישראליים.

המלצות למדיניות

התועלות האסטרטגיות והיישומיות של הבינה המלאכותית עשויות לשרת גורמים אנטי-ישראליים בדיוק כפי שהם עשויים לעמוד לצד ישראל. על כן מומלץ כי הגורמים הממשלתיים במערך ההסברה יפעלו בהקדם להטמעת השימוש בבינה מלאכותית במערך הדיפלומטיה הציבורית, בהתחשב באתגרים הכלולים בשימוש כזה. ארבע דרכי הפעולה הבאות עשויות להיות תוכנית ראשונית לשם כך:

- **מתן תמיכה ממשלתית לפיתוח פתרונות לשיפור יכולות ה-ChatGPT:** כפי שתואר לעיל, היכולות הקיימות ב-ChatGPT עשויות לספק ביצועים משופרים למשימות בתחום ההשפעה, אך הן עלולות לכלול כשלים פוטנציאליים משמעותיים. בהתאם לכך מומלץ כי המגזר הציבורי יישם פתרונות שיעודדו חברות טכנולוגיה להציג פתרונות יעילים וחדשניים לאותן בעיות. אפשר לעשות זאת באמצעות האקטונים (אירועי יזמות אינטנסיביים, שבמסגרתם מפתחים המשתתפים פתרונות טכנולוגיים לאתגרים המוגדרים על ידי מארגני האירועים) כגון זה **שקיימה** ממשלת קנדה ב-2023. פרסום קולות קוראים ממשלתיים למחקרים בתחום הבינה המלאכותית, דוגמת זה **שפורסם** באוגוסט 2022, עשוי להיות פתרון מסוג נוסף. יש לוודא כי קולות קוראים עתידיים יתמקדו בחקר מאפייניו וחסרונותיו הייחודיים של ה-ChatGPT, הרלוונטיים לתחום הדיפלומטיה הציבורית. יישום

של פעילות חברתית לגיטימית בעיני קהלים זרים. ככל שמושגים כאלו יופיעו בתוצרים המופקים מה-ChatGPT, עלולים גורמי הסברה ישראלים לקדם באופן בלתי מודע נרטיבים אנטי-ישראליים.

סוגיה זו מתקשרת לבעיה רחבה יותר – בעיית ההטיה המובנית **הקיימת במודלי שפה גדולים** (large language models). מודלים אלו, לרבות ChatGPT, מאומנים לנתח כמויות גדולות של טקסטים ולהפיק על בסיסם חומרים חדשים. בתהליך יצירת המודלים המפתחים **מזינים** להם כמויות גדולות של מידע ממקורות מגוונים, כמו ספרים ואתרי אינטרנט, המשמשים לאימון המודלים לביצוע משימתם. לעיתים התוצרים המופקים מהם עלולים לשקף תפיסות פוגעניות ובלתי מבוססות על סמך מגדר, גזע או קבוצה חברתית, וכך השימוש בבינה מלאכותית **עלול להוביל** לקבלת החלטות בלתי אובייקטיבית, המנציחה דעות קדומות. בתחום מדיניות חוץ עלול ה-ChatGPT להמליץ על גיבוש מסרים פוגעניים המופנים לקהלי יעד זרים מסוימים, מה שכמובן עלול לפגוע בתדמיתה של ישראל. בהתאם לכך **ניתן** להשתמש ב-ChatGPT להפקה מהירה של טיוטות מסרים שונות, שאת ניסוחן יחדדו גורמי הסברה אנושיים.

אפיק הפעילות השלישי הוא שימוש ב-ChatGPT לבדיקת עובדות (fact-checking) בתכנים המתפרסמים בתקשורת. בדומה לשני אפיקי הפעילות הקודמים, גם כאן הממצאים לגבי ביצועיו של ה-ChatGPT מעורבים: ה-ChatGPT **מסוגל** אומנם ברוב המקרים לסווג נכונה משפטים נכונים וכוזבים, אולם ביצועיו אינם עקביים, כך שטענה מסוימת שתיבדק פעמיים או תנוסח בשתי צורות שונות **תדורג** על ידו בשני אופנים שונים. בהתאם לכך ארגוני בדיקת עובדות טוענים כי התועלת העיקרית של שימוש בבינה מלאכותית בעבודתם היא היכולת לזהות יותר טענות שיש לבדוק. הצורך הזה הפך למהותי יותר לא רק בשל העובדה שהשימוש בבינה מלאכותית עלול להגדיל את היקף המידע הכוזב, אלא גם מאחר שהמשאבים העומדים לרשות בודקי העובדות **מצומצמים במידה משמעותית** לעומת אלו של מפיצי מידע כוזב. הפרכה מהירה של שקרים באמצעות עובדות עשויה לסייע לישראל לצמצם את תפוצתם, ובכך להקטין את הנזק התדמיתי הפוטנציאלי של **אירועים שונים**.

באותו נושא יש לציין פיתוח טכנולוגי ישראלי בשם **SAVEE**, המאפשר לגולשים להגיב לתכנים של הכחשת השואה ברשתות באמצעות שימוש בבינה מלאכותית יוצרת (Generative AI) לניסוח תגובות נגדן. פיתוח יוזמות מעין אלו לבדיקת עובדות לא רק מאפשר להפיק תגובות מהירות יותר נגד מסרים אנטי-ישראליים ואנטישמיים, אלא גם מעניק לכל אדם החפץ בכך את

משרדי הממשלה עם הארגונים הלא־ממשלתיים, בשל הפער בנהלים ובחוקים אשר להם כפופות שתי קבוצות הארגונים המשתפות את המידע. בהתאם לכך מומלץ כי הטמעת השימוש בבינה מלאכותית בדיפלומטיה הציבורית תיתן מענה לאתגרים אלו, למשל על ידי חידוד הבעיות שיש לפתור, כדי לשפר את ההבנה באשר לסוג המידע שעל הארגונים המעורבים לשאת.

• **הטמעת השימוש ב־ChatGPT באופן נסיוני בעבודה השוטפת של מערך ההסברה:** לאחר הגדרת הקריטריונים לשימוש בבינה מלאכותית במסגרת ממשלתית ולאחר התגברות על אתגרי התיאום שצוינו, אפשר יהיה להשתמש ב־ChatGPT במסגרת פרויקטים ניסיוניים מצומצמים, שיעסקו בשיפור היבטים מובחנים של פעילות הדיפלומטיה הציבורית. לדוגמה, ניתן לנתח עמדות של קהלי יעד מוגדרים במדינת יעד לתקופה מוגבלת, ולבחון את טיב ביצועי הבינה המלאכותית. בה בעת ניתן לשלב שימוש בתרחישים מבוססי בינה מלאכותית כחלק מהכשרת צוערים במשרד החוץ, ולהשוות את ביצועי החניכים שיעברו אותם לאלו של חניכים שלא התנסו בתרחישים אלו.

אופיר בראל הוא חוקר בסדנת יובל נאמן למדע, טכנולוגיה וביטחון באוניברסיטת תל אביב. מחקרו מתמקד בניתוח מבצעי השפעה המתנהלים ברשתות החברתיות. ophir.barel@gmail.com

הפתרונות הללו עשוי להוביל לפיתוח שיפורים טכנולוגיים רבים, בשל **מספרן הגדול** של חברות ההזנק הישראליות בתחום הבינה המלאכותית היוצרת.

• **הגדרת עקרונות מנחים לשימוש מבוקר ב־ChatGPT:** בשל בעיות אי־הדיוק וההטיה הקיימות בבינה מלאכותית, מומלץ כי משרדי הממשלה יקבעו עקרונות שיבטיחו את מיצוי היתרונות הפוטנציאליים, תוך צמצום הופעתן והשפעתן של הבעיות. לצורך כך ניתן למשל לקבוע כי יותר השימוש רק בצ'טבוטים שעומדים בקריטריונים הכלולים בחוק הבינה המלאכותית של האיחוד האירופי (EU AI Act), **המצוי בשלבי אישור** של מוסדות האיחוד האירופי. על פי החוק חברות המפתחות צ'טבוטים, דוגמת OpenAI וגוגל, יצטרכו למלא **שורה של דרישות** רגולטוריות מחמירות, כגון הבטחת יכולתם של מפעילים אנושיים להשפיע על אופן פעילותם, בטרם ישווקו באופן מסחרי.

• **מתן מענה לאתגרים ביורוקרטיים לשימוש בין־ארגוני בבינה מלאכותית:** כפי שצוין קודם לכן, פעולות הדיפלומטיה הציבורית מערבות גופים רבים, בעיקר אלו המשתייכים למגזר הציבורי. על פי **המחקר הקיים**, אימוץ בינה מלאכותית בשיתופי פעולה בין־משרדיים במגזר הציבורי עלול להיתקל בקשיים הקשורים לשיתוף המידע הנחוץ לשימוש בה (למשל בשל הבדלים במדיניות הפרטיות בין הארגונים). ההבדלים עלולים להיות חריפים יותר בעת עבודת