יוני 2025
סיון תשפ"ה

המלצות להתנהלות הציבור בשימוש אישי במערכות בינה מלאכותית יוצרת

רקע

מערכות בינה מלאכותית יוצרת (Generative AI) הן מערכות ממוחשבות שמטרתן "ליצור באמצעות שאילתה קצרה תוכן חדש - טקסטים, תמונות, סרטונים, מוזיקה, קוד מחשב ועוד, המופק בהתבסס על מידע אימון (training data) שהוזן למערכות".[1] בין הכלים הנפוצים בתחום ניתן לציין מערכות כמו ChatGPT, Claude, Dall-E, Midjourney, Gemini ואחרים.

הסיכונים והאתגרים הנובעים מפיתוח מערכות בינה מלאכותית ומהשימוש בהן כוללים נושאים רבים,[2] ובהם פגיעה בפרטיות. השימוש במערכות בינה מלאכותית יוצרת מעצים סיכון זה, בהיותו מבוסס על מידע והוראות (פרומפטים) המוזן על ידי המשתמש במהלך השימוש. הזנת המידע עשויה ליצור סיכון משמעותי לפרטיות ומהווה אתגר להגנה על המידע האישי.

מסמך זה סוקר חלק מן הסיכונים לפרטיות בשימוש במערכות בינה מלאכותית יוצרת למטרות אישיות ביתיות **בלבד** ומציע המלצות להתנהלות הציבור שיאפשרו להקטין סיכון זה.

יובהר, כי מכלול היבטי הפרטיות הנובעים מפיתוח מערכות בינה מלאכותית ומשימוש בהן, ויישום הוראות חוק הגנת הפרטיות על מערכות אלה מפורט בהרחבה בטיוטת הנחייה שפרסמה הרשות להגנת הפרטיות לא מכבר בנושא "תחולת חוק הגנת הפרטיות על מערכות בינה מלאכותית".[3]

לפיכך, ההמלצות שלהלן אינן עוסקות בשימוש בבינה מלאכותית לצרכי עבודה או למטרה עסקית או ארגונית, ואינן גורעות מן החובות והדרישות החוקיות החלות על שימוש במערכות בינה מלאכותית למטרות אלה, כמפורט בהנחיית הרשות. ההמלצות גם לא עוסקות בדרישות דינים ספציפיים נוספים הנוגעים לפרטיות ולסודיות של שימוש במידע אישי של אחרים, כגון בהזנת מידע ומסמכים של לקוחות ומיוצגים של משרד עורכי דין או מידע ומסמכים של מטופלים על ידי רופאים ופסיכולוגים, או בעניין פגיעה כללית בפרטיותם של אחרים בניגוד להוראות פרק א' לחוק הגנת הפרטיות. עוד יובהר כי המסמך אינו מבקש למנוע את השימוש במערכות בינה מלאכותית יוצרת, אלא לתת כלים להתנהלות נכונה של המשתמשים בהיבטי ההגנה על פרטיותם ברשת.

---

[1] בינה מלאכותית יוצרת: הזדמנויות, סיכונים ורגולציה, מרכז המידע והמחקר של הכנסת, 2023.
[2] להרחבה: עקרונות מדיניות, רגולציה ואתיקה בתחום הבינה המלאכותית, משרד החדשנות, המדע והטכנולוגיה, 2023.
[3] הנחיית הרשות להגנת הפרטיות בנושא "תחולת הוראות חוק הגנת הפרטיות על מערכות בינה מלאכותית" (אפריל 2025), שפורסמה כטיוטה להערות הציבור. ההנחיה זמינה כאן.

<u>מידע אישי בשימוש במערכות בינה מלאכותית יוצרת</u>

הרצון בהשגת תוצרים איכותיים ומותאמים לצרכים של המשתמש באמצעות שימוש במערכות בינה מלאכותית מזמין במקרים רבים הזנה של מידע אישי, לעיתים רגיש ובהיקף ניכר. השימוש במידע אישי במערכות אלו נעשה בהתאם למדיניות הפרטיות של כל מערכת ולתכליות המצוינות בה, שיכולות לכלול אחסון של המידע לתקופות שונות, שימוש בו לצורך אימון המערכת, העברה לצד שלישי וכל תכלית אחרת.

עם הרחבת השימוש במערכות בינה מלאכותית יוצרת פורסמו המלצות לדרכים שיסייעו בהפקה של תוצרים מותאמים להקשר, לתרחיש או לאוכלוסייה מסוימת ("הנדסת פרומפטים"). בין אלו ניתן למצוא המלצות לכתוב פרומפטים באופן פשוט, ישיר ומפורט,[4] או לתת למערכת "פרסונה", כלומר להורות למערכת לענות מתוך פרספקטיבה או תפקיד.[5]

הרצון להגיע לתוצרים יעילים בעזרת מערכות בינה מלאכותית יוצרת עשוי להוות אתגר בכל הנוגע לשמירה על המידע האישי, מאחר **שהזנה של מידע אישי כהוראה, הנחיה, שאלה או שאילתה למערכות בינה מלאכותית עשויה להוביל לפגיעה בפרטיות.** מלבד חשיפת המידע האישי עצמו, השאילתה עשויה לאפשר הסקה של מידע אישי נוסף מתוך שילוב והצלבת נתונים עם מקורות אחרים.

מידע אישי: דוגמה להוראה (פרומפט) ומה ניתן להסיק ממנה
פרומפט לדוגמה: "אני אוהב לבשל ומתעניין בריצה למרחקים ארוכים. אשמח לקבל המלצות על ציוד ספורט ועל מתכונים קלים להכנה". **מה ניתן להסיק מפרומפט זה?** **מצב בריאותי**: עיסוק בריצה עשוי להצביע על אורח חיים פעיל, ובשילוב עם מקורות מידע חיצוניים – על הסתברות או פוטנציאל לבעיות אורתופדיות או צורך בשיקום.

פרומפטים שנראים תמימים כשלעצמם מאפשרים יצירת פרופיל מקיף על תחומי עניין, הרגלים ופרטים אישיים שונים אודות האדם, בוודאי כאשר ישנו רצף של מספר פרומפטים.

[4] המלצות להנדסת פרומפטים בשימוש במערכת Claude: קישור
[5] המלצות להנדסת פרומפטים בשימוש במערכת ChatGPT: קישור

מידע אישי: דוגמה לרצף של הוראות (פרומפטים) ומה ניתן להסיק משילובן
פרומפט 1: "אני מחפשת המלצות על מסעדות טבעוניות טובות באזור תל אביב".
פרומפט 2: "אני מתעניינת באירועים למפתחי תוכנה שמתקיימים בחודש הבא".
פרומפט 3: "איזה שעון חכם מומלץ לאנשים שרצים באופן קבוע?".
מה ניתן להסיק מפרומפטים אלו?
מיקום: צפי לשהיה באזור תל אביב בזמן הקרוב.
אורח חיים: פעילה ובריאה, טבעונית, רצה באופן קבוע.

המידע המוזן בפרומפט אינו המידע היחיד הנחשף בשימוש במערכות בינה מלאכותית יוצרת. נתונים נוספים נובעים משמירת היסטוריה של הפעילות, חשבונות משתמש, מיקום ההתחברות, מערכת ההפעלה ונתונים אחרים.[6] מידע נוסף זה מאפשר להסיק פרטים נוספים על אדם מתוך צירוף הנתונים, גם אם כל פרט בפני עצמו אינו מאפשר זיהוי.

<u>סיכונים למידע האישי המוזן למערכות בינה מלאכותית יוצרת</u>

1. <u>חשיפת המידע האישי בשימוש במערכת</u>: מידע אישי שהוזן על ידי המשתמש ונשמר במערכת הבינה המלאכותית או משמש לאימון ושיפור האלגוריתם עשוי להשפיע על התוצרים המופקים עבור משתמשים אחרים במערכת. לדוגמה שמכם, כתובתכם, מספר הטלפון או כתובת הדוא"ל שלכם המוזנים למערכת בינה מלאכותית, עלולים להופיע ברשימה או בתוצר שהמערכת תתבקש להפיק עבור משתמש אחר.
2. <u>העברת מידע אישי ממערכת ה-AI לצד שלישי וסיכוני אבטחת מידע</u>: העברת המידע האישי לצד שלישי יכול להיות מנוצל לדיוג (Phishing),[7] לשימוש במידע כדי ליצור אמון עם המשתמש, לשילוב עם נתונים מרשתות חברתיות ולחשיפת פרטים על חברים ומשפחה. העברת המידע האישי לצד שלישי עלול לשמש גם להתחזות ומשלוח הודעות המכילות פרטים אישיים אמיתיים וקישורים כוזבים, להתאמה של קמפיינים פוליטיים או חברתיים ומגוון שימושים נוספים.

כמובן שבדומה לכל מערכת דיגיטלית, במידה ותוקף מקבל גישה לחשבון המשתמש או למערכת הבינה המלאכותית עצמה, הוא עשוי לקבל גישה, בין היתר, להיסטוריית ההוראות (פרומפטים)

[6] דוגמה – פירוט הנתונים הנאספים (וביניהם נתוני מערכת ההפעלה, כתובת IP, מיקום ונתוני שימוש) כפי שמצוין במדיניות הפרטיות של Gemini: <u>ק.ישור</u>.
[7] דיוג (Phishing) הוא ניסיון התחזות או הונאה שמטרתו להשיג מידע אישי.

ולפיכך לקבל מידע אישי רב אודות המשתמש. שילוב מידע זה עם מקורות גלויים נוספים עשוי לאפשר הסקה של מידע נוסף אודות המשתמש.

<u>המלצות להתנהלות הציבור בשימוש במערכות בינה מלאכותית יוצרת</u>

1. <u>מומלץ לצמצם, ככל האפשר, הזנה של מידע אישי</u> שאינו נדרש לשימוש במערכות בינה מלאכותית יוצרת. כתיבה כללית מצד אחד, וממוקדת בשאלה עצמה מצד שני, מפחיתה את הסיכון לחשיפת מידע אישי, ועדיין מאפשרת השגת תוצאות רלוונטיות מהמערכת.
2. <u>מומלץ להימנע, ככל האפשר, מאזכור פרטים מזהים</u>, כגון שם, גיל, מיקום, מצב משפחתי או מקצוע אם הם אינם נחוצים והכרחיים לשאלה. השתדלו במיוחד שלא לשתף מידע פיננסי (כמו מספרי חשבון בנק או מספרי כרטיסי אשראי), מידע רפואי (כולל פרטי ביטוח רפואי) וכל מידע אישי רגיש אחר. מומלץ בהחלט שלא לחשוף גם מידע אישי רגיש על אחרים, כגון על ילדיכם.
3. <u>מומלץ לנסח שאלות כלליות</u>: במקום "אני מחפשת", השתמשו ב"מהם", "אילו", או "איד". התמקדו במידע הנדרש: <u>השמיטו פרטים מיותרים שאינם קריטיים לתשובה</u> (לדוגמה, מספר הילדים או המיקום).

דוגמאות לניסוח הוראות (פרומפטים) עם וללא מידע אישי
פרומפט: "אני אמא לשלושה ילדים קטנים, גרה בירושלים ומחפשת רעיונות להפעלות בסוף השבוע לילדים בגילאי 4-8".
ניסוח ללא מידע אישי: "תציג לי בבקשה רעיונות להפעלות בסוף השבוע שמתאימות לילדים באזור ירושלים בגילאי 4-8" (הוסר המידע על מספר הילדים והמיקום המדויק של המשתמשת).
פרומפט: "אני מתכנת בן 35 מחיפה, שמחפש המלצות על אירועים בתחום ה-AI שיתקיימו בקרוב".
ניסוח ללא מידע אישי: "אילו אירועים בתחום ה-AI צפויים להתקיים בקרוב באזור חיפה?" (הוסר המידע על המקצוע, הגיל והמיקום המדויק של המשתמש).

4. <u>מומלץ להשתמש בהתממה (אנונימיזציה) באופן שמטשטש פרטים אישיים מזהים</u>, אך עדיין מאפשר לקבל תוצר מותאם, מדויק ורלוונטי. התממה יכולה להתבצע באמצעות הכללה (כגון הורדת רמת דיוק של נתון או קבוצת נתונים), או שינוי ערכים אקראי (הוספת "רעש") על מנת למנוע זיהוי של המשתמש.

דוגמה לניסוח הוראה (פרומפט) תוך שימוש בהתממה
פרומפט: "אני אבא לשלושה ילדים בני 6, 9 ו-12, גר בירושלים, ומחפש המלצות על בתי ספר שבהם יש דגש על מתמטיקה". **הכללה על ידי המרת גיל לטווח ("גילאי יסודי" במקום גיל מדויק):** "אילו בתי ספר עם דגש על מתמטיקה מתאימים לילדים בגילאי יסודי וחטיבת ביניים בירושלים?"

5. מומלץ לבקש מהמערכת למחוק את כלל המידע האישי או פרטים מסוימים מפעם לפעם במידה ואפשרות זו קיימת. במרבית כלי הבינה המלאכותית היוצרת קיימת אפשרות לקבל את המידע האישי הקיים בכלי אודות המשתמש, או לצייר/לתאר את המשתמש לפי הידוע למערכת. מומלץ להשתמש באפשרות זו כדי לדעת את היקף וסוג המידע האישי הנשמר במערכת על אודות המשתמש.
6. מומלץ לשקול לבטל את האפשרות של אימון מערכת הבינה המלאכותית היוצרת על מידע המוזן אליה, אם אפשרות כזו קיימת, כדי לצמצם את הסיכון לפרטיות מהמידע הנשמר במערכת. בדקו האם נאספים נתוני מיקום כחלק מהשימוש במערכת, ושקלו לבטל את האיסוף, אם אפשרות כזו קיימת. בדקו האם ישנה סקירה אנושית של חלק מהשיחות כחלק מתהליכי אבטחת האיכות או הפיתוח של המערכת ושקלו לבטל סקירה זו, אם אפשרות כזו קיימת. באופן כללי, תנאי השימוש ומדיניות הפרטיות של מערכות בינה מלאכותית יוצרת משתנים מעת לעת, ומומלץ לעקוב אחר השינויים, בפרט במערכות העיקריות שבהן אתם משתמשים, כדי להכיר את אופן העיבוד של המידע האישי שלכם, ולערוך שינויים מתאימים בהגדרות הפרטיות כדי להגן על פרטיותכם ברשת.