כ״ה בשבט תשפ״ה
23 בפברואר 2025

מדריך לטכנולוגיות מגבירות-פרטיות

הקדמה

טכנולוגיות מגבירות-פרטיות (Privacy-Enhancing Technologies - PETs) הן אוסף של גישות ופתרונות דיגיטליים המסייעים להגנה על מידע אישי. השימוש בטכנולוגיות מגבירות-פרטיות מאפשר להפיק את הערך הנדרש מהמידע תוך שמירה על הפרטיות והגנה על המידע האישי באמצעות (1) ערפול המידע האישי הנדרש לשימוש וצמצום רמת הפירוט שלו, (2) הקטנת הסיכון לחשיפת המידע האישי במהלך העיבוד (3) והגברת השליטה בשימוש במידע האישי.

העיסוק בתחומי טכנולוגיות מגבירות-פרטיות לאורך השנים האחרונות משותף לרשויות הגנת מידע במדינות שונות[1]. השימוש בטכנולוגיות מגבירות-פרטיות כחלק מעיצוב מערכות במרחב הדיגיטלי תורם לעמידתן בדיני הגנת הפרטיות ואבטחת המידע, לצד מילוי המחויבות הבסיסית להגן על פרטיות המידע וכתנאי מאפשר ליצירת אמון מול המשתמשים.

מטרות המסמך

1. הצגת סקירה של טכנולוגיות מגבירות-פרטיות נפוצות ומתפתחות.
2. הצגת דוגמאות ושיקולים לשילוב טכנולוגיות מגבירות-פרטיות כחלק מהמענה להגנת הפרטיות בתהליכים, מערכות ופרויקטים.

היקף המסמך

המסמך מתמקד במספר טכנולוגיות מייצגות ובולטות מתוך מגוון טכנולוגיות מגבירות-פרטיות הזמינות כיום. המסמך מתאר את עקרון הפעולה הבסיסי של כל טכנולוגיה ומציין שיקולים מרכזיים בקבלת ההחלטה על שימוש בטכנולוגיה מסוימת כחלק ממכלול האמצעים להגנה על המידע האישי.

המסמך אינו מפרט טכנולוגיות כלליות בתחומי אבטחת מידע. כמו כן, המסמך אינו עוסק בהרחבה בנורמות המשפטיות (חוקים ותקנות) או בתהליכים ארגוניים (כגון תסקיר השפעה על הפרטיות) המשפיעים על שילוב טכנולוגיות מגבירות-פרטיות.

טכנולוגיות מגבירות-פרטיות ועיצוב לפרטיות (Privacy by Design)

עקרון העיצוב לפרטיות (Privacy By Design) הוא שילוב היבטי הגנת הפרטיות ואבטחת מידע כחלק מהתכנון של מערכות ותהליכים, החל משלב הדרישות וקונספט הארכיטקטורה ולאורך כל מחזור החיים של אותם מערכות ותהליכים. עקרון זה אומץ בהחלטות שונות לאורך השנים[2] ונכנס

[1] פירוט נוסף מופיע בסעיף 'מקורות מידע מרכזיים' בהמשך פרק זה ובהפניות בגוף המסמך.
[2] לדוגמה, החלטה בנושא עיצוב לפרטיות – הכינוס הבינלאומי ה-32 של נציבי הגנת הפרטיות שהתקיים בירושלים בשנת 2010.

ppa@justice.gov.il | 073-3928529 | *3103
קרית הממשלה, ת.ד. 33503, תל-אביב 6133401 | www.go.gov.il/ppa

כחלק מדרישות רגולציית הגנת המידע של האיחוד האירופי (GDPR)[3], ומאז אף נכנס כחלק מההגדרות בחוקים ותקנות במדינות שונות.

השימוש בטכנולוגיות מגבירות-פרטיות הוא חלק מהאמצעים להגנה אופטימאלית על המידע האישי כחלק מיישום עקרון העיצוב לפרטיות בתהליכי איסוף מידע, שמירתו והגנה על הגישה אליו, תוך שיש בהם כדי להרחיב באופן משמעותי את אפשרויות השימוש במידע אישי, לרבות מידע אישי בעל רגישות מיוחדת בהיקפים גדולים במיוחד. ההחלטה על הטמעה של טכנולוגיות מגבירות-פרטיות יכולה להתקבל כתוצאה של עריכת תסקיר השפעה על פרטיות (Privacy Impact Assessment)[4] בטרם תחילת עיבוד המידע, או כחלק מניהול סיכונים מתמשך לאורך חיי המערכת.

טכנולוגיות מגבירות-פרטיות משלימות תהליכים ואמצעים משפטיים וארגוניים הזמינים למפתחי המערכות והכלים הטכנולוגיים ומסייעות ליישום מיטבי של הוראות דיני הגנת הפרטיות ושימוש הוגן במידע אישי.

אופן הצגת הטכנולוגיות

תיאור הטכנולוגיות במסמך כולל הסבר המתאר את העקרון העומד ביסוד הטכנולוגיות השונות ואופן פעולתן, במידת העניין בצירוף דוגמה או איור. בהמשך מובאים שיקולים ביישום הטכנולוגיה ודוגמאות ליישומים קיימים במגוון תחומים ועולמות תוכן. ניתנת התייחסות לאתגרים ומגבלות ביישום הטכנולוגיות לצד סיכונים ונושאים הדורשים תשומת לב מיוחדת.

הדוגמאות המובאות במסמך זה מדגימות את עקרון ואופן הפעולה של טכנולוגיות מגבירות-פרטיות בצמצום ופישוט ניכר לצורך המחשה וללא פירוט כלל הפרטים והתהליכים הטכניים המתחייבים ביישום מעשי. הדוגמאות מציגות שימושים נפוצים או בולטים מעולמות תוכן מגוונים ואין בכך כדי להמליץ על דרך פעולה אחת או להגביל את השימושים האפשריים. לשם יישום מעשי של טכנולוגיות מגבירות-פרטיות נדרשת התייחסות מעמיקה למגוון רב של נושאים ותחומים תוך התחשבות במאפיינים הייחודיים בכל מקרה ופרויקט.

לאורך המסמך נכללים קישורים לפרסומים של חוקרים ואנשי תעשיה מישראל ומהעולם להרחבה והעמקה, לצד פרסומים של גופי ממשל בעולם. מקורות אלו נועדו לאפשר הרחבה והעמקה במידע המוצג, לתת דוגמאות ולאפשר עיון נוסף בהקשר לנושאים המתוארים במסמך.

אוכלוסיית יעד

המסמך מיועד לשמש את האחראים והממונים על הערכת סיכוני הפרטיות ויישום המענה המתאים בפרויקטי פיתוח מערכות ושירותים דיגיטליים. לא פחות מכך, המסמך עשוי לשרת את אנשי הפיתוח בעולמות הדיגיטל בהטמעה של טכנולוגיות מגבירות-פרטיות החל משלב הייזום בפרויקט ולאורך מחזור חייו. באופן פרטני, המסמך עשוי לשמש את בעלי התפקידים הבאים:

1. ממוני הגנת הפרטיות (Data Protection Officers – DPOs) ויועצים משפטיים העוסקים בתחומי הגנת הפרטיות.

---

[3] Art. 25 GDPR - Data protection by design and by default.

[4] לפירוט נוסף ראו מדריך עזר מתודולוגי לעריכת תסקיר השפעה על הפרטיות שפורסם על ידי הרשות (2022).

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 קרית הממשלה, ת.ד. 33503, תל-אביב

2. מנהלי מוצר ומנהלי פרויקטי פיתוח, הטמעה והפעלה של מערכות מידע, שירותים ומוצרים דיגיטליים.

השימוש במסמך אינו דורש ידע טכני או רקע טכנולוגי. כפועל יוצא מכך, רמת הפירוט בהתייחס לכל טכנולוגיה מאפשרת להכיר את מהותה ולהעריך את מידת התאמתה לתחום או משימה ספציפיים. פרטים נוספים לצורך מימוש ומענה לאתגרים וסיכונים יימצאו במקורות נוספים, ולצורך כך ניתן להיעזר בקישורים להרחבה המובאים בסיום של כל פרק במסמך זה. יש לציין שלצורך בחינת המוצר הטכנולוגי המתאים אין בכך כדי להחליף אנשי מקצוע המומחים בתחום זה.

מקורות מידע מרכזיים

המסמך מתבסס בחלקו על חומרים בנושאי טכנולוגיות מגבירות-פרטיות שפורסמו על ידי ה-OECD (Emerging privacy-emerging technologies), האו"ם (The PET Guide – The United Nations Guide on Privacy-Enhancing Technologies for Official Statistics) ורשויות הגנת פרטיות בעולם. חלק מההפניות במסמך לקוחות מתוך דוח מקיף שפרסמה רשות הגנת המידע של בריטניה (ICO - Information Commissioner's Office) Privacy-enhancing technologies (PETs). חלק אחר מהחומרים מופיע במאגר פרסומים ממקורות רשמיים בנושאי טכנולוגיות מגבירות-פרטיות באתר של Future of Privacy Forum (FPF). חומרים נוספים שעליהם מבוססים סעיפים ספציפיים של המסמך מצוינים בהערות שוליים ובגוף הטקסט.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 קרית הממשלה, ת.ד. 33503, תל-אביב

תוכן העניינים

הקדמה ..... - 1 -
מטרות המסמך ..... - 1 -
היקף המסמך ..... - 1 -
טכנולוגיות מגבירות-פרטיות ועיצוב לפרטיות (Privacy by Design) ..... - 1 -
אופן הצגת הטכנולוגיות ..... - 2 -
אוכלוסיית יעד ..... - 2 -
מקורות מידע מרכזיים ..... - 3 -
מבוא לטכנולוגיות מגבירות-פרטיות ..... - 5 -
טכנולוגיות מגבירות-פרטיות – הגדרה ..... - 5 -
טכנולוגיות מגבירות-פרטיות – קטגוריות ..... - 6 -
מיפוי טכנולוגיות מגבירות-פרטיות מרכזיות ..... - 7 -
טכנולוגיות המסייעות לערפול המידע האישי הנדרש לשימוש וצמצום רמת הפירוט שלו ... - 9 -
התממה – Anonymization ..... - 9 -
מידע סינתטי – Synthetic Data ..... - 16 -
פרטיות דיפרנציאלית – Differential Privacy ..... - 17 -
טכנולוגיות המסייעות להקטנת חשיפת מידע אישי במהלך השימוש ..... - 21 -
הצפנה הומומורפית – Homomorphic Encryption ..... - 21 -
הוכחה באפס ידיעה – Zero Knowledge Proof ..... - 23 -
חישוב רב-משתתפים – Multi-Party Computation ..... - 24 -
הצלבת מערכי מידע – Private Set Intersection ..... - 26 -
למידה מבוזרת – Federated Learning ..... - 27 -
סביבת ביצוע מהימנה – Trusted Execution Environment ..... - 29 -
טכנולוגיות המסייעות לשליטה במידע האישי ..... - 31 -
מאגרי מידע בשליטת נושא המידע – Personal Data Stores ..... - 31 -
כלי תיעוד ושקיפות – Documentation and Transparency Tools ..... - 32 -
סיכום ..... - 33 -

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 קרית הממשלה, ת.ד. 33503, תל-אביב

מבוא לטכנולוגיות מגבירות-פרטיות

טכנולוגיות מגבירות-פרטיות הן חלק ממכלול האמצעים המיועדים לסייע לצמצום סיכונים להגנת הפרטיות ולאבטחת מידע אישי. שילוב טכנולוגיות מגבירות-פרטיות הוא חלק מארגז הכלים של ממונה הגנת הפרטיות (Data Protection Officer - DPO) בארגון, לצד כלים נוספים כגון תסקיר השפעה על פרטיות (Privacy Impact Assessment)[5]. טכנולוגיות מגבירות-פרטיות מהוות אלמנט משלים לכלים משפטיים וארגוניים קיימים כחלק מתהליכי בקרה ומעקב אחר סיכוני הגנת הפרטיות ואבטחת המידע.

טכנולוגיות מגבירות-פרטיות – הגדרה

לאורך השנים ניתנו מגוון הגדרות לטכנולוגיות מגבירות-פרטיות[6]. במסמך זה נגדיר טכנולוגיות מגבירות-פרטיות באופן הבא:

טכנולוגיות מגבירות-פרטיות הן אוסף של שיטות, תהליכים וכלים דיגיטליים המסייעים בהגנה על מידע אישי. טכנולוגיות מגבירות-פרטיות מאפשרות לערפל[7] את המידע האישי ולצמצם את רמת הפירוט שלו, להקטין את הסיכון לחשיפת מידע אישי במהלך העיבוד ולהגביר את השליטה בשימוש במידע אישי.

כדי לעמוד על ערכן וחשיבותן של טכנולוגיות מגבירות-פרטיות עבור שימוש במידע, נבקש להפנות להגדרת "מידע אישי" בחוק הגנת הפרטיות, התשמ"א-1981 – ההגדרה הבסיסית ביותר של החוק – אשר עודכנה מהיסוד במסגרת תיקון מס' 13 לחוק, ובנוסחה החדש קובעת כך: ""מידע אישי" - נתון הנוגע לאדם מזוהה או לאדם הניתן לזיהוי; לעניין הגדרה זו, "אדם הניתן לזיהוי" - מי שניתן לזהותו במאמץ סביר, במישרין או בעקיפין, בכלל זה באמצעות פרט מזהה, כגון שם, מספר זהות, מזהה ביומטרי, נתוני מיקום, מזהה מקוון, או נתון אחד או יותר הנוגע למצבו הפיזי, הבריאותי, הכלכלי, החברתי או התרבותי".[8] **המשמעות היא שכאשר טכנולוגיה מגבירת-פרטיות הופכת מידע לכזה שלא ניתן לקשור במאמץ סביר לאדם מזוהה, במישרין או בעקיפין, אף לא באמצעות זיהוי חוזר, המגבלות הקבועות בחוק הגנת הפרטיות ותקנותיו לא יחולו על המידע, והשימוש בו כלל לא יהווה פגיעה בפרטיות**. למען הסר ספק נבהיר, כי עצם עיבוד המידע האישי המזוהה בתוך הארגון, לרבות תוך שימוש בטכנולוגיות מגבירות-פרטיות, כפוף כמובן להוראות חוק הגנת הפרטיות.

---

[5] הפניה להרחבה – בהערת שוליים 4.

[6] סקירה בנושא מופיעה בדוח של ה-OECD בנושא טכנולוגיות מגבירות-פרטיות, סעיף 2.2.

[7] ערפול מידע – בהתאמה למונח "Data obfuscation" על פי ה-OECD.

[8] יצוין, כי חוק הגנת הפרטיות (תיקון מס' 13), התשפ"ד-2024, המהווה רפורמה מקיפה בחוק, עתיד להיכנס לתוקפו ביום 14.8.2025.

ppa@justice.gov.il | 073-3928529 | *3103
קרית הממשלה, ת.ד. 33503, תל-אביב 6133401 | www.go.gov.il/ppa

יצוין, כי השימוש בטכנולוגיות מגבירות-פרטיות, כגון אנונימיזציה (התממה), מסייע בהגנה על המידע האישי אך חשוף לסיכונים שונים היוצרים קושי משמעותי להגיע למצב כאמור, בו הנתונים לא ייחשבו עוד מידע אישי כמשמעותו בחוק הגנת הפרטיות.

טכנולוגיות מגבירות-פרטיות – קטגוריות

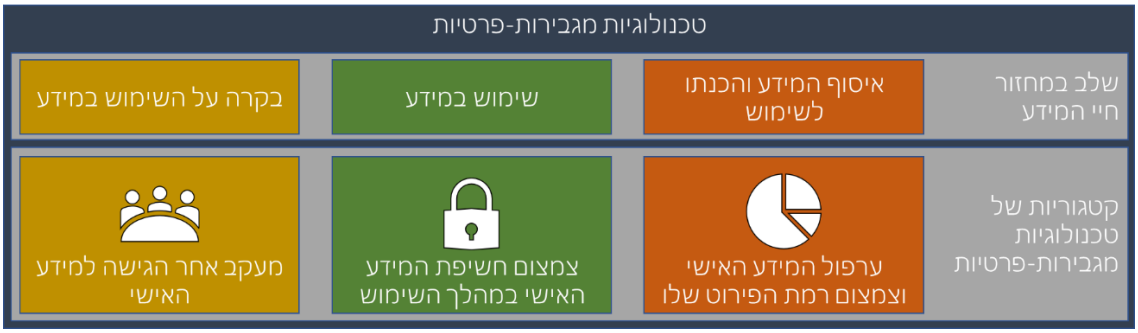
טכנולוגיות מגבירות-פרטיות הן משפחה מגוונת של שיטות, תהליכים וכלים דיגיטליים המתאימים לשלבים שונים במחזור חיי המידע, ובייחוד לשלושה תחומים רחבים:

1. איסוף המידע והכנתו לשימוש.
2. שימוש במידע.
3. בקרה על השימוש במידע.

טכנולוגיות מגבירות-פרטיות נבדלות ביניהן באופן הפעולה וצורת המימוש, אך ניתן לחלק אותן לשלוש קטגוריות רחבות בהתאם לעקרון פעולתן:

1. ערפול המידע האישי הנדרש לשימוש וצמצום רמת הפירוט שלו.
2. צמצום חשיפת המידע האישי במהלך השימוש.
3. מעקב אחר הגישה למידע האישי.

החלוקה לפי שלבים במחזור חיי המידע ועקרון הפעולה מייצרת שלוש קטגוריות עיקריות של טכנולוגיות מגבירות-פרטיות, בהתאם לאיור 1:



איור 1: קטגוריות עיקריות של טכנולוגיות מגבירות-פרטיות

להלן פירוט של הקטגוריות העיקריות של טכנולוגיות מגבירות-פרטיות:

1. **איסוף המידע והכנתו לשימוש**: ערפול המידע האישי הנדרש לשימוש או צמצום רמת הפירוט שלו, בין היתר על ידי הסרת פרטים מזהים, שינוי הנתונים, טשטוש הערכים המדויקים או הוספת "רעש"[9]. דוגמאות לטכנולוגיות בתחום זה הן התממה ומידע סינתטי.

[9] הוספת "רעש" למידע דומה להוספת קולות רבים להקלטה כדי למנוע זיהוי של המידע בשיחה. הוספת "רעש" היא הוספת נתונים אקראיים או שינוי אקראי של חלק מהנתונים כדי להקשות על זיהוי נתוני המקור. כאשר הוספת "רעש" נעשית באופן המתאים, שימושיות הנתונים עשויה להישמר לצד הגנה טובה יותר על המידע.

ppa@justice.gov.il | 073-3928529 | *3103
קרית הממשלה, ת.ד. 33503, תל-אביב 6133401 | www.go.gov.il/ppa

2. **שימוש במידע**: צמצום החשיפה של המידע האישי במהלך העיבוד ואף שימוש במידע ללא צפיה בו במהלך העיבוד. דוגמאות לטכנולוגיות בתחום זה הן הצפנה הומומורפית וחישוב רב-משתתפים.
3. **בקרה על השימוש במידע**: הגדרה של כללים והרשאות לגישה למידע אישי והצגת הנתונים הנוגעים לזהותו של מי שניגש למידע, לסוג המידע ולמועד הגישה. דוגמאות בתחום זה הם מאגרי מידע בשליטת נושאי המידע[10] וכלי תיעוד ושקיפות.

מיפוי טכנולוגיות מגבירות-פרטיות מרכזיות

הטבלה להלן מפרטת את הטכנולוגיות מגבירות-הפרטיות בהתאם לקטגוריות שהוגדרו בחלק הקודם. לכל טכנולוגיה בטבלה מתוארים בקצרה: עקרון הפעולה; דוגמאות ליישומים; ואתגרים ומגבלות. פירוט נוסף נמצא בפרק המתאים לכל טכנולוגיה בהמשך מסמך זה:

קטגוריה	טכנולוגיה	עקרון פעולה	דוגמאות ליישומים	אתגרים ומגבלות
טכנולוגיות המסייעות **לערפול המידע האישי** וצמצום רמת הפירוט שלו	**התממה** (Anonymization)	הסרה וטשטוש של מאפיינים מזהים במידע	השמטת ערכים, הכללה ומיסוך נתונים	○ איבוד מידע והורדת רזולוציה ○ נאמנות למידע המקורי, הטיות (bias) ושימושיות ○ ידע טכנולוגי בהפעלת הכלים
	מידע סינתטי (Synthetic Data)	הפקת מאגרי מידע חדשים עם קשר סטטיסטי למידע המקור	הפקת מידע לאימון מודלי בינה מלאכותית	
	פרטיות דיפרנציאלית (Differential Privacy)	הוספת רעש אקראי למידע	הנגשת מאגרי מידע לצרכי מחקר סטטיסטי	
טכנולוגיות המסייעות **לצמצום חשיפת המידע האישי**	**הצפנה הומומורפית** (Homomorphic Encryption)	ביצוע חישובים על מידע בהיותו מוצפן	ביצוע חישובים על מידע רגיש	○ עומס חישובי ○ ידע טכנולוגי במימוש
	הוכחה באפס ידיעה (Zero Knowledge Proof)	הוכחת נכונות מידע ללא חשיפת המידע עצמו	אימות תכונות (כגון גיל, ניסיון)	○ יישומים מצומצמים וייעודיים

[10] כלומר, האדם עצמו: "נושא המידע" – האדם שעל אודותיו קיים מידע במאגר המידע (ההגדרה קבועה בתקנה 1 לתקנות הגנת הפרטיות (אבטחת מידע), תשע"ז-2017).

ppa@justice.gov.il | 073-3928529 | *3103
קרית הממשלה, ת.ד. 33503, תל-אביב 6133401 | www.go.gov.il/ppa

קטגוריה	טכנולוגיה	עקרון פעולה	דוגמאות ליישומים	אתגרים ומגבלות
במהלך השימוש	**חישוב רב-משתתפים** (Multi-Party Computation)	עיבוד נתונים משותף מבלי שאף צד יחלוק את הנתונים שלו עם הצדדים האחרים	ביצוע חישובים הדורשים איסוף מידע רגיש ממספר משתתפים	○ תעבורת רשת ○ עומס חישובי ○ ידע טכנולוגי במימוש
	הצלבת מערכי מידע (Private Set Intersection)	מציאת נתונים המשותפים בין מאגרי מידע ללא חשיפת נתונים שאינם משותפים	○ איתור אנשי קשר משותפים ○ ניטור מגע בין אנשים	
	למידה מבוזרת (Federated Learning)	למידת מכונה מבוזרת תוך מזעור הנתונים המשותפים בין הצדדים	אימון מודלים ללמידת מכונה	○ דיוק התוצאות ○ ידע טכנולוגי במימוש
	סביבת ביצוע מהימנה (Trusted Execution Environment)	עיבוד נתונים בחלק מבודד ומאובטח של מערכת המחשב	○ תשלום מאובטח ○ אימות זהות באמצעים ביומטריים	○ עומס חישובי ○ ידע בחומרה ותוכנה ○ אמון ביצרן/מפתח של הסביבה
טכנולוגיות המסייעות **למעקב אחר הגישה למידע האישי**	**מאגרי מידע בשליטת נושאי המידע** (Personal Data Stores)	ניהול הרשאות גישה למידע אישי ותיעוד הגישות בפועל למידע האישי	אחסון מידע אישי וניהול הרשאות גישה אליו	אחריות של בעל המאגר לניהול המידע, הרשאות הגישה ואבטחת המידע
	כלי תיעוד ושקיפות (Documentation and Transparency Tools)	תיעוד מדויק ומלא של הגישה למידע האישי תוך מתן אפשרות מעקב לנושא המידע	הצגת פירוט אירועי גישה למידע אישי המוחזק במאגרי מידע	הטמעה בארגונים

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 קרית הממשלה, ת.ד. 33503, תל-אביב

ראשית כאמור, מידע שאינו מזוהה ולא ניתן לקשור אותו במאמץ סביר לאדם מזוהה, אף לא באמצעות זיהוי חוזר, לא ייחשב "מידע אישי" המוגן על ידי חוק הגנת הפרטיות, בהתאם להגדרה החדשה אשר נחקקה בתיקון מס' 13 לחוק.

שנית, עקרון צמצום המידע האישי הוא אחד מעקרונות דיני הגנת הפרטיות[11], ולפיו יש לאסוף ולשמור אך ורק את המידע המינימלי הנדרש וההכרחי למטרת האיסוף או למטרת מאגר המידע, וזאת מבחינת היקף המידע הנשמר, סוג המידע, זמן שמירתו וכדומה.

טכנולוגיות מגבירות-פרטיות מאפשרות לערפל את המידע האישי הנדרש לשימוש או לצמצם את רמת הפירוט שלו וייתכן שיוכלו במקרים מסוימים להפוך מידע אישי למידע שאינו מזוהה, תוך עמידה במטרת האיסוף או מטרת המאגר. מעבר לאמור, צמצום רמת פירוט המידע הרלוונטי והיקפו הוא חלק מהותי בתהליכי עיצוב לפרטיות (Privacy by Design) ומסייע בהקטנת סיכוני אבטחת מידע.

התממה – Anonymization

התממה היא הסרת מאפיינים או שינוי ערכים על מנת לצמצם או למנוע זיהוי של נושא המידע. דרכים נפוצות להתממה הן הסרת שדות הכוללים מזהים ישירים (שדות המאפשרים זיהוי של אדם באופן ישיר, כגון שמות, מספרי תעודת זהות או אחד הפרטים המזהים הנכללים בהגדרת "מידע אישי" בתיקון מס' 13 לחוק הגנת הפרטיות), הסרת שדות עם מזהים עקיפים (שדות שבשילוב עם מידע נוסף מאפשרים זיהוי של אדם, כגון מקצועות ומקומות עבודה), הכללה (כגון הורדת רמת דיוק של נתון או קבוצת נתונים), או הוספת רעש אקראי למידע. צמצום רמת פירוט המידע הרלוונטי והיקפו מגבירים את ההגנה על הפרטיות, אך עשויים לפגוע בערך ובשימושיות של המידע ליישומים שונים.

פירוט: אפשרויות נפוצות ודוגמאות להתממה
1. **השמטת ערך – Attribute Suppression**: מחיקת ערך (עמודה מטבלה) שאינו נדרש למטרת עיבוד המידע. לדוגמה, אם בסיס הנתונים כולל את שם התלמיד, שם המורה וציון – לצורך חישוב התפלגות הציונים לפי מורה ניתן למחוק את שם התלמיד ולעבוד רק עם שם המורה והציון. 2. **השמטת רשומה – Record Suppression**: מחיקת רשומה (שורה מטבלה) ייחודית או חריגה הניתנת לזיהוי חוזר בקלות יחסית. לדוגמה, אם בסיס הנתונים כולל נתוני הכנסה לפי גיל, ויש רק תושב אחד בקבוצת הגיל מעל 100, הכנסתו ניתנת לזיהוי וניתן למחוק את הרישום אודותיו, בהנחה שהדבר לא יפגע באופן מהותי במדגם.

[11] ראו טיוטת מסמך המדיניות של הרשות בנושא צמצום מידע (Data Minimization) (2021).

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 קרית הממשלה, ת.ד. 33503, תל-אביב

3. **מיסוך נתונים חלקי – Character Masking**: מיסוך נתונים תיאוריים על ידי החלפת חלק מהתיאור בתו קבוע. לדוגמה, שינוי הנתון בשדה מיקוד מ-96554 ל-96XXX מוריד את הרזולוציה של המידע כך שלא ניתן לזהות אזור מגורים ספציפי, אך בשיטות מיקוד מסוימות משאיר את היכולת לזהות עיר או אזור כללי.
4. **פסאודונימיזציה – Pseudonymization**: החלפת מידע מזהה במידע "פיקטיבי", לדוגמה החלפת מספר זהות במחרוזת אקראית. פסבדונימים יכולים להיות בלתי הפיכים או הפיכים (על ידי שימוש בטבלאות המרה[12] או במנגנוני ההצפנה). ניתן להשתמש בפסבדונים לצורך זיהוי וקישור של מידע אישי בין מאגרי מידע שונים מבלי לחשוף את זהות הפרט (למשל, לראות את היסטוריית האשפוזים של אדם במספר בתי חולים).
5. **הכללה – Generalization**: הורדת רמת דיוק של נתון או קבוצת נתונים[13]. לדוגמה, החלפת גובה ההכנסה המדויק (15,980 ש"ח) בטווח הכנסה (10-20 אלף ש"ח), או שימוש בערך "משך האשפוז בימים" במקום "תאריך תחילת אשפוז" ו-"תאריך סיום אשפוז".
6. **ערבוב ערכים – Shuffling**: החלפת ערכי עמודות בין הרשומות במאגר המידע (כאשר אין צורך לחקור או להתייחס לתלויות בתוך כל רשומה ורשומה). לדוגמה, החלפה אקראית של הערכים בעמודת משקל גוף בין אנשים שונים.
7. **הוספת רעש – Noise addition**: שיבוש הערכים אשר עשויים לתרום לזיהוי בהצלבה למקורות מידע נוספים / חיצוניים. ניתן להוסיף רעש אקראי או לעגל ערכים. לדוגמה, החלפת ערכי סכומים של 18.1 ש"ח ו-44.9 ש"ח ב-18 ש"ח ו-45 ש"ח בהתאמה.
8. **מידע סיכומי – Data Aggregation**: החלפת מידע במדדים סטנדרטיים כגון סכומים, ממוצעים או סטיית תקן. לדוגמה, במקום סכומי תרומות ותאריכים מדויקים – מספר התרומות שהתקבלו בכל חודש והסכום שלהן.

דוגמאות ליישום שיטות נפוצות להתממה מובאות בעמוד הבא.

[12] לצד צמצום המידע האישי באמצעות פסאודונימיזציה, שמירה של טבלאות המרה מגדילה את כמות הטבלאות במאגר המידע. השימוש בטבלת המרה יכול להעלות את הסיכון של אחזור המידע האישי וקישור בין המידע האישי למידע שעבר פסאודונימיזציה במקרה של אירוע אבטחת מידע.

[13] הכללה היא אפקטיבית בתנאי שקיימת קבוצה מספיק גדולה של פרטים או לחלופין שההכללה לא מאפשרת לזהות קבוצת פרטים בתוך אוכלוסייה גדולה.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 קרית הממשלה, ת.ד. 33503, תל-אביב

שיטה	תיאור	לפני התממה	אחרי התממה
השמטת ערך	מחיקת ערך (עמודה מטבלה)	משה: זכר דנה: נקבה	זכר נקבה
השמטת רשומה	מחיקת רשומה (שורה מטבלה)	משה: זכר דנה: נקבה	דנה: נקבה
מיסוך נתונים חלקי	החלפת חלק מהתיאור בתו קבוע	משה: moshe@a.com דנה: dana@b.com	משה: *****@a.com דנה: ****@b.com
פסאודונימיזציה	החלפת מידע מזהה במידע "פיקטיבי"	משה: זכר דנה: נקבה	זכר: A12c19V0 נקבה: LVV098C
הכללה	הורדת רמת דיוק של נתון או קבוצת נתונים	משה: 15,000 ש"ח דנה: 22,000 ש"ח	משה: 10-20 אש"ח דנה: 20-30 אש"ח
ערבוב ערכים	החלפת ערכי עמודות בין הרשומות	משה: 68 ק"ג דנה: 59 ק"ג	משה: 59 ק"ג דנה: 68 ק"ג
הוספת רעש	שיבוש ערכים באמצעות רעש או עיגול המידע	משה: 68 ק"ג דנה: 59 ק"ג	משה: 70 ק"ג דנה: 55 ק"ג
מידע סיכומי	החלפת מידע במדד סיכומי (כגון סכום או ממוצע לאורך זמן)	שם: תאריך: סכום משה: 12.10.23: 50 ש"ח דנה: 11.6.23: 13 ש"ח	שם: שנה: סה"כ משה: 2023: 170 ש"ח דנה: 2024: 180 ש"ח

סיכונים מרכזיים לפרטיות בשימוש במאגרי מידע שעברו התממה כוללים חשיפת זהות (בידוד נושא מידע ספציפי משאר נושאי המידע) וחשיפת מאפיין (קבלת ידיעה על תכונה של הפרט):

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 קרית הממשלה, ת.ד. 33503, תל-אביב

1. **חשיפת זהות (Identity disclosure)** – זיהוי מלא של רשומת הפרט בתוך המאגר המידע. לרוב זיהוי מלא מתאפשר עבור רשומות בעלות מאפיינים ייחודיים. הכרעה בשאלה מה הסבירות לחשיפת זהות ובידוד נושא מידע תלויה במאפייני המידע במאגר ומספר קטגוריות המאפשרות בידוד נושא המידע בו.
2. **חשיפת מאפיין (Attribute disclosure)** – שיוך תכונה מסוימת מתוך מאגר המידע לפרט מסוים. לדוגמה, בהינתן קובץ נוכחות אנונימי שלפיו כל עובדי מחלקה מסוימת הגיעו ביום מסוים למקום העבודה אחרי השעה 10 בבוקר, ניתן לקבוע לגבי פרט בודד כי הוא איחר לעבודה באותו יום רק על סמך המידע שהוא עובד באותה מחלקה, גם אם זהותו לא מופיעה מפורשות בקובץ.

חשוב לציין שטכניקות התממה נפוצות, כגון השמטה, מיסוך או הכללה, אינן חסינות מפני זיהוי חוזר (Re-Identification) מלא או חלקי. לאורך השנים פורסמו מקרים רבים שבהם התאפשר זיהוי תכונה או הסקת מסקנות אודות אדם מתוך מידע שעבר התממה, למשל על ידי הצלבה עם מאגר מידע אחר.

אתגר ההבחנה בין מידע אישי למידע אנונימי הוא משמעותי ונדון בהרחבה, בין היתר, בדוח הצוות הבינמשרדי בנושא בינה מלאכותית בסקטור הפיננסי[14]. לפי דוח זה, לפעולת ההתממה יתרונות משמעותיים בהגנה על הפרטיות של נושאי המידע, אלא שאין בכך כדי להבטיח הגנה מוחלטת על מידע אישי ועל פרטיותם של נושאי המידע, בוודאי בעידן של בינה מלאכותית ויכולותיה, המקלות על ביצוע זיהוי חוזר של נושאי מידע מתוך מידע מותמם.

דוגמה: זיהוי חוזר של מידע שעבר התממה – פרס נטפליקס (Netflix Prize)

בשנת 2006 חברת נטפליקס שחררה מאגר מידע אנונימי של דירוגי סרטים במסגרת תחרות לחיזוי דירוג סרטים על בסיס דירוגי עבר, ללא כל מידע נוסף על המשתמשים. המאגר לא כלל כל מידע על המשתמשים מלבד מזהה ייחודי (פסבדונים) לציון מספר דירוגים של אותו משתמש. הרשומות כללו את הנתונים הבאים:

1. מזהה משתמש.
2. פרטי הסרט.
3. תאריך הדירוג.
4. דירוג הסרט (1-5).

בשנת 2007 שני חוקרים מאוניברסיטת טקסס דיווחו, כי הצליחו לזהות פרטי משתמשים ייחודיים על ידי התאמה של נתוני הדירוג של נטפליקס לנתוני הדירוג בבסיס נתונים פתוח באתר

[14] הצוות הבין-משרדי לבחינת השימוש בבינה מלאכותית בסקטור הפיננסי - בינה מלאכותית בסקטור הפיננסי, דוח ביניים להערות הציבור, אוקטובר 2024.

ppa@justice.gov.il | 073-3928529 | *3103
קרית הממשלה, ת.ד. 33503, תל-אביב 6133401 | www.go.gov.il/ppa

של IMDB - Internet Movie Database. התאמת הפרטים בוצעה על בסיס הדמיון בתאריכי הדירוג לאותם סרטים בשני בסיסי הנתונים. רמת ההתאמה שדווחה בחלק מהמקרים הייתה גבוהה ביותר.

חומרת הפגיעה בפרטיות במקרה זה נבעה מכך, שדירוגים בנטפליקס ניתנו בעילום שם, וחלק מהמשתמשים התייחסו לסרטים שמהם אפשר היה להסיק מסקנות על דעות פוליטיות, אמונות דתיות ופרטים אישיים נוספים. קישור הדירוגים ב-IMDB אפשר קבלת פרטים מזהים של המדרגים ובהתאם לכך להסיק נתונים רגישים אודותם.

דוגמה: הצלבת מידע לזיהוי חוזר של מידע שעבר התממה – מושל מסצ'וסטס

בשנות ה-90 פורסמו למטרות מחקר נתונים אנונימיים על ביקורים בבתי חולים של עובדי ציבור במדינת מסצ'וסטס בארה"ב. מהמידע הוסרו כלל המזהים הישירים, כגון שמות, כתובות ומספרי זיהוי (SSN - Social Security number).

החוקרת לטניה סוויני (Latanya Sweeney) רכשה את פנקס הבוחרים בעיר קיימברידג', שבה התגורר מושל מסצ'וסטס באותה עת. היא הצליבה את הנתונים בשני המאגרים וגילתה, שרק שישה אנשים בקיימברידג' חלקו את יום ההולדת של מושל מסצ'וסטס, מחציתם היו גברים ורק אחד מהם גר באזור המיקוד של המושל. כך באמצעות שימוש במידע גלוי היא הצליחה לקשר מידע רפואי רגיש לפרטים של אדם ספציפי. תרשים החפיפה בין המאגרים במחקר מובא להלן:

במחקר המשך[15], אותה חוקרת הראתה, שכ-87% מאוכלוסיית ארה"ב ניתנים לזיהוי חוזר בסבירות גבוהה רק על בסיס ידיעת תאריך לידה, מיקוד ומין.

[15] L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 קרית הממשלה, ת.ד. 33503, תל-אביב

הסיכונים לזיהוי או הסקת מסקנות מתוך מידע שעבר התממה קפצו מדרגה בעידן נתוני העתק (Big Data), נוכח נגישותם של מאגרי מידע רבים וההתפתחות הטכנולוגית בתחום. בהתאם לכך, התממה מספקת של נתונים והערכת הסיכון לזיהוי חוזר מהווים אתגר משמעותי.

גישה אחת להבטחת התממה של הנתונים מופיעה בחוק הניידות והאחריות של ביטוח בריאות האמריקאי (HIPAA[16]), בסעיף 164.514 לחוק, ומתבססת על שני ערוצים אפשריים:

1. **הערכת מומחה** להתממה של הנתונים: קביעה של אדם בעל ידע וניסיון מתאימים המפעיל עקרונות ושיטות סטטיסטיים ומדעיים מקובלים כי הסיכון לכך שהמידע יכול לשמש, לבד או בשילוב עם מידע זמין סביר אחר, לזיהוי נושא המידע הוא קטן מאוד, תוך תיעוד שיטות ותוצאות הניתוח המצדיקות זאת.
2. **הורדת מאפיינים** להבטחת התממה (גישת נמל מבטחים – Safe Harbor): הורדה מהמידע של 18 משפחות מזהים ישירים ועקיפים (כגון כתובות, לוחיות רישוי, תאריכים, מזהי גלישה באינטרנט ועוד).

היכולת לחלץ מידע אישי מתוך מידע אנונימי תלויה במגוון גורמים. מודלים ומדדים לאנונימיות מידע מאפשרים להעריך את רמת הקושי בזיהוי מחדש ולהשוות בין רמת הפירוט של בסיסי נתונים שונים.

מודל k-Anonymity הוא אחד המודלים המסייעים להערכת רמת הסיכון לזיהוי חוזר של מידע במאגר נתונים. מאגר הייחוס כולל מזהים ישירים (שדות שמאפשרים זיהוי של אדם – כגון שם או כתובת מייל), מזהים עקיפים (שדות שבשילוב עם מידע נוסף מאפשרים זיהוי של אדם – כגון תאריך לידה או שכונת מגורים) ומאפיין רגיש (מידע שאין רוצים לחשוף את זהות האדם שאליו הוא שייך). ביחס למאגר זה, ובהנחה שנמחקו ממנו כלל המזהים הישירים שיטת k-Anonimity בוחנת את המספר הקטן ביותר של רשומות עם מזהים עקיפים זהים.

במאגר המקיים הגדרות של k-Anonymity, לכל רשומה יש לפחות k רשומות הזהות לה במזהים עקיפים (בהנחה שהמאגר אינו כולל מזהים ישירים). בכך מתאפשרת הגנה טובה יותר מפני התקפות קישור מאפיינים, מאחר שאין אפשרות לקשר את המאפיין הרגיש לרשומה מסוימת מתוך k הרשומות ללא מידע נוסף.

[16]: נוסח ההוראה: https://www.govinfo.gov/content/pkg/CFR-2017-title45-vol1/pdf/CFR-2017-title45-vol1-sec164-514.pdf. תיאור נוסף ניתן למצוא ב- Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, US Department of Health and Human Services, Office for Civil Rights, 2010 http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/guidance.html#_edn32.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 קרית הממשלה, ת.ד. 33503, תל-אביב

דוגמה: מודל k-Anonymity

הטבלה להלן מתארת חלק ממאגר רשומות הכוללות מזהים עקיפים של גיל, מין ומצב משפחתי, יחד עם מאפיין רגיש של מספר ילדים. במאגר זה – קבוצה של 3 גברים ששלושתם נשואים ובטווח גילאים בין 40-50 מקיימת את התנאי של אנונימיות עם k=3. קבוצה זו מסומנת במסגרת עבה בטבלה להלן.

מזהים עקיפים			מזהה רגיש
גיל	מין	מצב משפחתי	מספר ילדים
[40-50]	ז	נשוי	2
[40-50]	ז	נשוי	3
[40-50]	ז	נשוי	4
[30-40]	נ	נשואה	3
[30-40]	נ	נשואה	2
[30-40]	נ	נשואה	3

אם, כמו בטבלה לעיל, גם במאגר כולו עבור כל צירוף ייחודי של מזהים עקיפים (גיל, מין ומצב משפחתי) יש לפחות 3 רשומות זהות – אפשר לומר שכלל המאגר מקיים את התנאי של אנונימיות עם k=3 ביחס למספר הילדים. המשמעות היא שלא ניתן לבודד רשומה שתהיה ייחודית מבחינת המזהים העקיפים, אלא רק 3 רשומות ומעלה.

ככל שערך k עולה, כך יורדת הסבירות לזיהוי חוזר, אך גם רמת ההכללה עולה והדבר עשוי להפחית ממידת השימושיות של הנתונים. בנוסף לכך, מודל k-Anonymity מניח כי לכל פרט יש רשומה אחת במאגר. אם הנחה זו לא מתקיימת (למשל, עבור מאגר נתוני אשפוז שיכולים לכלול מספר רשומות עבור אותו אדם), נדרש להעלות את ערך k באופן משמעותי, כך שיקטן הסיכוי שכל הערכים בקבוצת k או רובם יהיו שייכים לאותו אדם.

טווח הערכים ל-k-Anonymity נקבע בהתאם למאפייני המידע בכל מקרה לגופו מאחר שהסיכוי לזיהוי חוזר משתנה ממקרה למקרה ותלוי בפרמטרים רבים – אין אפשרות לקבוע פרמטר כללי מסוים, והגדרתו נדרשת להיות בהתאם להערכת סיכוני פרטיות עבור כל מקרה, תהליך ומערכת.

הגדרת המנגנון והפרמטרים להתממה צריכה להיעשות כחלק מתפיסה כללית של עיצוב לפרטיות (Privacy by Design) וכחלק מתהליכים ארגוניים שלמים הכוללים הערכה מתמשכת ותסקיר השפעה על הפרטיות (Privacy Impact Assessment). יש לשוב ולבחון מנגנון ופרמטרים אלו מחדש בכל שינוי של התנאים הרלוונטיים בארגון ובסביבתו, כמו גם מעת לעת.

ppa@justice.gov.il | 073-3928529 | *3103
קרית הממשלה, ת.ד. 33503, תל-אביב 6133401 | www.go.gov.il/ppa

התממה – מקורות להרחבה
1. מדריך להתממה של רשות הגנת המידע בסינגפור (PDPC): Guide to Basic Anonymization 2. מתווה להערכת סיכוני זיהוי חוזר של רשות הגנת המידע האוסטרלית: The de-identification decision-making framework 3. מדריך להסרת פרטים מזהים ממידע מובנה של רשות הגנת המידע של אונטריו (קנדה): De-identification Guidelines for Structured Data 4. מדריך בנושא פסאודונימיזציה שיצא להערות הציבור על ידי הרשות האירופית להגנת מידע (EDPB): Guidelines 01/2025 on Pseudonymisation 5. מאמרם של טל ז'רסקי ושרון בר-זיו העוסק בהרחבה בהיבטי התממה: פרטיות במשבר זהות: אסטרטגיות הסדרה בעידן ההתממה

מידע סינתטי – Synthetic Data

מידע סינתטי הוא נתונים המיוצרים בדרך כלל לפי דפוסים ומאפיינים סטטיסטיים של נתונים אמיתיים (המכילים מידע אישי). מידע סינתטי נועד לאפשר להפיק תוצאות דומות לעיבוד מידע אישי, ללא שימוש במידע אישי אמיתי. מעבר להגנה על הפרטיות, מידע סינתטי עשוי לשמש למגוון מטרות נוספות, ביניהן ככלי ליצירת מערכי מידע גדולים לאימון מודלים של בינה מלאכותית או הפקת מידע הכולל מגוון רב של מקרי קצה ואירועים נדירים. השימוש במידע סינתטי לעיתים משולב יחד עם מידע שאינו סינתטי (אמיתי) ולעיתים מחליף את המידע האמיתי לחלוטין.

דרכים נפוצות להפקת מידע סינתטי:

1. **הפקה פרמטרית**: בניית מודל של המאפיינים הסטטיסטיים של מידע המקור (לדוגמה – ממוצע, סטיית תקן, קורלציה וכו') והפקת נתונים סינתטיים באמצעות שימוש במודל.
2. **למידת מכונה ובינה מלאכותית**: הפקת מידע סינתטי מתוך תבניות ומאפיינים שהופקו מתוך למידה מנתוני האימון.

דוגמה: מידע סינתטי – מאגר מידע של הבנק העולמי[17]
בשנת 2023 הבנק העולמי פרסם מאגר נתונים סינתטי למטרות סימולציה והדרכה. המאגר כולל מידע על 8,000 משקי בית, המייצגים מדגם של האוכלוסייה של מדינה דמיונית בעלת הכנסה בינונית. מאגר המידע כולל משתנים הנאספים בדרך כלל במפקדי אוכלוסין (כגון השכלה, משלח יד, מאפייני דיור ומאפיינים משפחתיים) ובסקרי משקי בית (כגון הוצאות משק בית או בעלות על נכסים).

[17] World Bank. (2023). Synthetic Data for an Imaginary Country, Sample, 2023 [Data set]. World Bank, Development Data Group.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 קרית הממשלה, ת.ד. 33503, תל-אביב

מאגר המידע נוצר באמצעות בינה מלאכותית מנתונים שעברו דגימה וקידוד מחדש באופן שאינו מאפשר לקשר בין המידע הסינתטי למידע האמיתי. מערך הנתונים נוצר למטרת הדרכה וסימולציה ואינו מיועד לייצג מדינה ספציפית כלשהי.

הרשומות במאגר מידע סינתטי אינן אמורות לכלול מידע אמיתי של אדם ספציפי, אלא רק לעיתים נדירות ובמקרה (למשל, כתוצאה מהגרלה אקראית). עם זאת, במקרים מסוימים ניתן יהיה להסיק מהמידע הסינתטי מידע אישי מזוהה. המתקפות העיקריות על מידע סינתטי:

1. **מתקפת שחזור** (Database Reconstruction) – זיהוי נתונים אישיים ששימשו ליצירת נתונים סינתטיים.
2. **מתקפת הסקת מאפיין** (Attribute Inference) – זיהוי מאפיין אישי כחלק ממידע ששימש ליצירת נתונים סינתטיים.
3. **מתקפת שייכות** (Membership Inference) – זיהוי מאפיינים של אדם או קבוצת אנשים ששימשו ליצירת הנתונים הסינתטיים.

השימוש במידע סינתטי מקטין מטבע הדברים את הסיכון לפגיעה בפרטיות ומאפשר שימושים נוספים במידע. עם זאת, בתנאים מסוימים, ייתכנו מצבים שבהם תהיה יכולת להגיע לזיהוי חוזר של מידע אישי (למשל, כאשר ישנן רשומות ייחודיות רבות). לכן גם בשימוש במידע סינתטי דורש ניתוח סיכוני פרטיות ונקיטת אמצעים להגנה על המידע (כגון היבלעות בהמון – k-Anonymity או פרטיות דיפרנציאלית).

מידע סינתטי – מקורות להרחבה
1. מדריך ליצירת מידע סינתטי של רשות הגנת המידע בסינגפור (PDPC): Proposed Guide on Synthetic Data Generation 2. מדריך למידע סינתטי של מכון אלן טיורינג וה-Royal Society (בריטניה): Synthetic Data - what, why and how? 3. מאמר בנושא מידע סינתטי של הארגון הבינלאומי למומחי פרטיות (IAPP): Synthetic data: What operational privacy professionals need to know

פרטיות דיפרנציאלית – Differential Privacy

פרטיות דיפרנציאלית היא גישה שפותחה עבור מאגרי נתונים הכוללים מידע אישי אך מיועדים למטרות עיבוד סטטיסטי (שאינו מיועד לחשוף מידע אישי). מהות הגישה של פרטיות דיפרנציאלית היא שהוספה, הסרה או שינוי של רשומה אחת במאגר המידע תשפיע במידה מועטה ביותר, אם בכלל, על תוצאות היישום של פונקציה סטטיסטית על מאגר המידע.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 קרית הממשלה, ת.ד. 33503, תל-אביב

במקום להסיר או לשנות נתונים כדי לטשטש מזהים, לצורך השגת פרטיות דיפרנציאלית יש להוסיף למידע האמיתי נתונים אקראיים, או רעש. המטרה היא להוסיף כמות מספקת של נתונים אקראיים כך שמידע אמיתי לא יהיה ניתן לזיהוי מתוך הרעש. פרטיות דיפרנציאלית עדיין מאפשרת לבצע ניתוח מדויק על נתונים מצטברים, מכיוון שלמרות הרעש הנוסף – הנתונים המשולבים יכולים לספק תוצאות מדויקות.

דוגמה: פרטיות דיפרנציאלית – מנגנון תשובה אקראית (Randomized response)

דוגמה להפעלת מנגנון המאפשר פרטיות דיפרנציאלית היא מנגנון תשובה אקראית (Randomized response). שיטה זו משמשת לאיסוף תשובות לשאלות אודות מידע רגיש ומאפשרת עיבוד מידע עבור קבוצה גדולה של אנשים מבלי שתהיה יכולת לקשר מענה ספציפי לאדם מסוים.

נניח שהנשאל צריך לענות על השאלה 'האם הצבעת בבחירות למפלגה X או למפלגה Y?'. הנשאל מטיל מטבע (מבלי שהשואל רואה), ופועל בהתאם לתוצאת ההטלה:

1. במקרה של עץ – משיב תשובה אקראית ("X" בהסתברות 50% ו"Y" בהסתברות 50%).
2. במקרה של פלי – עונה את התשובה האמיתית לשאלה.

במנגנון זה רק הנשאל יודע האם התשובה היא אמיתית או אקראית, ולכן לשואל אין אפשרות לדעת מה התשובה האמיתית של נשאל מסוים. עם זאת, המנגנון מאפשר להסיק מסקנות על אוכלוסיית המדגם, מאחר שבהנחה והמדגם הוא גדול מספיק – התפלגות התוצאות קרובה מאוד להתפלגות האמיתית.

להדגמת אופן חישוב התוצאה ללא ידיעת התשובות האמיתיות של הנשאלים – נניח שהתפלגות התוצאות (האמיתית) היא 60% מול 40% לטובת אחת המפלגות. בהפעלת מנגנון תשובה אקראית, עבור כמחצית מהנשאלים התפלגות התשובות תהיה אקראית (כלומר בערך 50:50), ועבור היתר – ההתפלגות תהיה אמיתית (60:40). לא נוכל לדעת מי מהנשאלים ענה אקראית ומי ענה תשובה אמיתית, אלא רק את ההתפלגות הכוללת (שהיא הממוצע בין שתי ההתפלגויות, כלומר בערך 45%:55%).

ממוצע זה מספיק כדי לחשב בעזרתו את ההתפלגות האמיתית, וניתן לעשות זאת על ידי הכפלת הפער (10%) פי שניים. כך נקבל שהפער האמיתי הוא 20%, כלומר ההתפלגות האמיתית תהיה בערך 60% מול 40%, כפי שהנחנו בדוגמה.

ppa@justice.gov.il | 073-3928529 | *3103
קרית הממשלה, ת.ד. 33503, תל-אביב 6133401 | www.go.gov.il/ppa

פרטיות דיפרנציאלית מבוססת על רעיון של "הפרש קטן" בין מידע המופק משתי קבוצות נתונים הנבדלות ברשומה אחת, כך שלא ניתן להסיק מידע אישי על אדם מסוים באמצעות השוואה בין התוצאות. קונספט זה מתאר הפקת מידע (תשאול) מקבוצת נתונים בשני מצבים:

1. כאשר נתון של אדם מסוים נמחק מקבוצת הנתונים.
2. כאשר הנתון של אותו אדם נשאר כפי שהוא בקבוצת הנתונים.

פרטיות דיפרנציאלית מאפשרת להבטיח כי התוצאה של התשאול תהיה דומה מאוד בשני המצבים לעיל. כלומר, ההשפעה של הוספת פרטי אדם תהיה זניחה, ועל כן הסיכוי לזליגת מידע אישי מתוך השוואת תוצאות העיבוד עם או בלי הנתונים אודות אותו אדם יהיה קטן מאוד.

דוגמה: מצב שבו פרטיות דיפרנציאלית לא נשמרת (עפ"י [Wood et al., 2018][18])

אוניברסיטה מפרסמת נתונים ממוצעים על ההכנסה של סטודנטים. לפי פרסומיה, בחודש אפריל למדו בה 304 סטודנטים, ומתוכם 30 השתכרו מעל 30 אלף ש"ח בחודש. בחודש מאי למדו בה 303 סטודנטים, ומתוכם 29 השתכרו מעל 30 אלף ש"ח בחודש.

אמנם אלו נתונים אגרגטיביים, אך ניתן להסיק מהם כי בחודש מאי עזב את האוניברסיטה סטודנט שהכנסתו היא מעל 30 אלף ש"ח בחודש. כפועל יוצא, חבריו לכיתה של הסטודנט שעזב יהיו מודעים לרמת הכנסתו.

דוגמה זו מתארת מצב שבו פרטיות דיפרנציאלית לא נשמרת, כלומר ניתן להסיק מידע מהשוואה בין מידע עבור קבוצות הנבדלות ביניהן בפרטיו של אדם אחד בלבד.

רכיב חשוב בפרטיות דיפרנציאלית הינו הערך אפסילון (ε) או "תקציב הפרטיות", אשר קובע מה תהיה רמת הרעש אשר יתווסף למערך הנתונים. הגדרה פורמלית של פרטיות דיפרנציאלית היא שההשפעה של הוספת או הסרת פרט בודד לא תביא לשינוי בערך השאילתה מעבר לפרמטר ε. ככל ש-ε קטן יותר, פרטיות המשתמשים גבוהה יותר, אך השימוש בכלי יחייב רעש חזק יותר ולכן המידע עשוי להיות פחות שימושי.

[18] A. Wood, M. Altman, A. Bembenek, M. Bun, M. Gaboardi, J. Honaker, K. Nissim, D. R. OBrien, T. Steinke, and S. Vadhan. "Differential privacy: A primer for a non-technical audience." Vanderbilt Journal of Entertainment & Technology Law 21, no. 1 (2018): 209-275.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 קרית הממשלה, ת.ד. 33503, תל-אביב

דוגמה: תקציב הפרטיות במנגנון פרטיות דיפרנציאלית (עפ"י [Wood et al., 2018][19])

איילת היא בעלת פוליסת ביטוח חיים על סך 100,000 ש"ח. חברת הביטוח קבעה עבורה פרמיה שנתית של 1,000 ש"ח, בהתבסס על טבלאות אקטואריות המעריכות כי לאשה בגילה של איילת סיכוי של 1% למות בשנה הבאה.

איילת בחרה להשתתף במחקר רפואי, שבו התגלה שיש לה סיכוי של 50% למות משבץ מוחי בשנה הבאה. אם מידע זה יעבור לחברת הביטוח – הפרמיה שלה תעלה ל-50,000 ש"ח ויותר.

לעומת זאת, בהנחה ונתוני המחקר ישוחררו באופן מצרפי, תוך הבטחת פרטיות דיפרנציאלית עם פרמטר $\varepsilon=0.01$, אזי יובטח כי ההערכה של חברת הביטוח לסיכוי של איילת למות בשנה הבאה תעלה לכל היותר ב: $1\%*(1+0.01)=1.01\%$. כלומר, הפרמיה של איילת תעלה לכל היותר ב-10 ש"ח.

פרטיות דיפרנציאלית היא כלי התלוי בנתונים, בסוג העיבוד ובמנגנון הפרסום של הנתונים. עבור נתונים בעלי שונות גבוהה יהיה צורך להוסיף רעש חזק יותר, כדי להגיע לרמה דומה של הגנה על המידע.

פרטיות דיפרנציאלית מאפשרת התממה כל עוד מתווספת רמת רעש נאותה, וכל עוד התשובה לשאילתה הינה מידע אנונימי (למשל מידע אגרגטיבי). התוצרים של פרטיות דיפרנציאלית יעילים לניתוח סטטיסטי של מגמות כלליות, אך מתאימים פחות לזיהוי אנומליות או דפוסים בתוך המידע עצמו עקב הרעש הנוסף. יש להדגיש כי תוצרי פרטיות דיפרנציאלית לא יניבו בהכרח מידע אנונימי, ועל כן יש לבחון כל מקרה לגופו. אם הכלי לא ייושם באופן נאות קיים חשש לדליפת מידע באמצעות שאילתות חוזרות ונשנות על ידי תוקף.

פרטיות דיפרנציאלית – מקורות להרחבה
1. מבוא לפרטיות דיפרנציאלית עבור אנשים ללא רקע טכני (אתר אוניברסיטת הרווארד): Differential privacy: A primer for a non-technical audience
2. פרויקט קוד פתוח של אוניברסיטת הרווארד לפיתוח כלים לפרטיות דיפרנציאלית: OpenDP
3. תיאור (מבוא) ומגוון מקורות בנושא פרטיות דיפרנציאלית באתר של ד"ר דמיאן דפונטיין (Dr. Damien Desfontaines): A friendly, non-technical introduction to differential privacy

[19] שם

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 קרית הממשלה, ת.ד. 33503, תל-אביב

טכנולוגיות המסייעות להקטנת חשיפת מידע אישי במהלך השימוש

הגנת מידע במנוחה (at rest) ובמעבר (in transit) היא חלק בלתי נפרד מתשתיות הדיגיטל המודרניות. טכנולוגיות מגבירות-פרטיות מרחיבות את הגנת המידע גם לנתונים במהלך השימוש (in use). טכנולוגיות אלו מאפשרות לצמצם את החשיפה של המידע ולהקטין את הסיכונים בהתאם באמצעות מגוון שיטות, תהליכים וכלים דיגיטליים.

הצפנה הומומורפית – Homomorphic Encryption

הומומורפי הוא מונח באלגברה שמשמעותו שווה צורה. הצפנה הומומורפית משתמשת במבנה אלגברי מיוחד המאפשר שימור של חלק מהתכונות של המידע המקורי במידע המוצפן. כך, לאחר ביצוע פעולות על המידע המוצפן (ללא פתיחתו) ניתן בתנאים מסוימים לחלץ תוצאה המתאימה לחישוב על מידע המקור. לפיכך, הצפנה הומומורפית מאפשרת שימוש במידע בהיותו מוצפן ללא צורך בפתיחת ההצפנה, ובהתאם מצמצמת את הסיכון בחשיפה של המידע לגורם בלתי מורשה במהלך העיבוד.

תהליך העבודה עם הצפנה הומומורפית מתחיל בהצפנה של המידע ויצירת מפתח הערכה (evaluation key) שמאפשר עבודה על המידע המוצפן. לאחר מכן מבוצע חישוב על המידע בצורתו המוצפנת (ללא חשיפתו). לבסוף התוצאה ניתנת לפתיחה באמצעות שימוש במפתח הסודי. בתהליך זה המידע מוגן (מוצפן) במהלך השימוש ועל כן מאפשר, בין היתר, ביצוע חישובים על מכשיר קצה מרוחק או לא מאובטח.

דוגמה: הצפנה הומומורפית – המחשת עקרון השימוש

הצפנה הומומורפית כוללת שימוש בפרוטוקולי הצפנה ייחודיים. כדי להמחיש את עיקרון השימוש בהצפנה הומומורפית נציג סכמה נאיבית (שאינה עושה שימוש במנגנוני הצפנה מודרניים מקובלים). בדוגמה נתייחס לשליחת שני מספרים לצורך ביצוע פעולת כפל בשרת חיצוני, כאשר אין מעוניינים שהשרת ייחשף למספרים עצמם או לתוצאת הכפל. ההצפנה שנפעיל בסכמה זו תהיה העלאה של המספרים בחזקת מספר סודי שישתנה עבור כל פעולת חישוב.

לצורך ההדגמה נבחר את המספרים 2 ו-3. נפעיל עליהם את ה'הצפנה' על ידי העלאתם בחזקת מספר סודי (לצורך הדוגמה – חזקת 2), כך שהמידע שיישלח לשרת יהיה המספרים 4 ו-9. השרת יבצע את פעולת הכפל ויחזיר את התוצאה – 36. על מספר זה אפשר יהיה לבצע את פעולת השורש וכך לקבל את התוצאה הנכונה – 6.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 קרית הממשלה, ת.ד. 33503, תל-אביב

בדוגמה זו המידע מוצפן לפני שהוא נשלח מהלקוח לעיבוד בשרת והפעולה מבוצעת על המידע בצורתו המוצפנת. התוצאה מועברת מהשרת ללקוח, שפותח את ההצפנה כדי לקבל את התוצאה, כך שבשום שלב המידע אינו נחשף לאף גורם מלבד הלקוח.

כאמור, סכמת מימוש זו משמשת להמחשה של עקרון הפעולה בלבד, ויישום הצפנה הומומורפית דורש מימוש של כלל האלגוריתמים ומנגנוני אבטחת המידע הרלוונטיים שפירוטם חורג מהיקף מסמך זה.

פרוטוקולי מימוש הצפנה הומומורפית מאפשרים מספר סוגים של יכולות:

1. **הצפנה הומומורפית חלקית** (Partially Homomorphic Encryption), **הצפנה הומומורפית מסוימת** (Somewhat Homomorphic Encryption), **הצפנה הומומורפית מרודדת** (Leveled Fully Homomorphic Encryption) – מאפשרות ביצוע חישובים תוך הגבלה על סוג הפעולות או מספר החישובים האפשרי.
2. **הצפנה הומומורפית מלאה** (Fully Homomorphic Encryption) – מאפשרת לבצע מספר בלתי מוגבל של חישובים ללא הגבלה על סוגי הפעולות.

הצפנה הומומורפית מייצרת עומס חישובי משמעותי המתחייב מעבודה על המידע כשהוא בצורתו המוצפנת בהשוואה לעיבוד מידע שאינו מוצפן. לכן, שיטה זו יעילה פחות כאשר יש צורך לעבד כמות גבוהה של מידע, ובהתאם לכך מומלץ למקד את השימוש בהצפנה הומומורפית למידע שהצפנתו תסייע באופן המירבי להורדת הרגישות והסיכון למידע האישי. התפתחויות טכנולוגיות אשר יתרמו לקצבי מחשוב גבוהים יותר ופיתוח אלגוריתמים יעילים יותר צפויים לתרום לצמצום המגבלות החישוביות במימוש הצפנה הומומורפית ובהתאם לכך להרחבת השימוש בטכנולוגיה זו.
השימוש בהצפנה הומומורפית דורש שימוש מדויק בכלים ותשתיות מתאימים. חשוב להקפיד על שימוש באלגוריתמים ותוכנות שנבדקו, רמות הצפנה (גודל מפתח) מתאימות ושמירה על המפתחות הפרטיים, כמו גם הגדרת תהליכי ייצור מפתח פרטי נוסף והצפנה מחדש של המידע שכבר הוצפן במקרה שהמפתח המקורי דלף או נחשף.

הצפנה הומומורפית – מקורות להרחבה
1. בלוג של OpenMined הכולל מידע נוסף על הצפנה הומומורפית ואופן מימושה: What is Homomorphic Encryption?
2. מדריך בנושא הצפנה הומומורפית של פרופ' בועז ברק (אוניברסיטת הרווארד): Fully homomorphic encryption: Introduction and bootstrapping

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 קרית הממשלה, ת.ד. 33503, תל-אביב

הוכחה באפס ידיעה – Zero Knowledge Proof

הוכחה באפס ידיעה היא טכניקה המאפשרת במקרים מסוימים ומול שאילתות מסוימות להוכיח תכונה או נכונות מידע ללא חשיפת המידע עצמו. הוכחה באפס ידיעה יכולה לאפשר להוכיח בגרות (מבלי לגלות תאריך לידה), מצב כלכלי (מבלי לגלות נתונים כלכליים), בעלות בנכס (מבלי לגלות פרטי עסקה), ובאופן עקרוני לתמוך בשיטות הזדהות כגון זיהוי פנים, טביעת אצבע ואישור קולי.
הוכחה באפס ידיעה ממומשת על ידי בקשה לביצוע סדרת פעולות שתוצאתן תלויה במידע מסוים. אם בכל פעם התוצאה שיציג מבקש האימות תהיה נכונה – ניתן יהיה לומר ברמת ביטחון גבוהה שמבקש האימות יודע את המידע הדרוש. כך ניתן להוכיח ידיעה של מידע מסוים מבלי שהמידע עצמו עבר בין הצדדים.

דוגמה: הוכחה באפס ידיעה – מימוש אפשרי

דוגמה למימוש הוכחה באפס ידיעה היא בקשה לקבלת ערך של פונקציה עבור קלט מסוים. אם המשתמש יחזיר ערך נכון של הפונקציה באופן עקבי מספר רב של פעמים, זה יעיד על סבירות גבוהה שיש לו ידע על הפונקציה, מאחר שהסבירות לניחוש מדויק של הערך יורדת עם כל סבב נוסף של בקשות.

בדוגמה באיור 2 הלקוח (איור ימני) מבקש מהשרת את ערכי הפונקציה הסודית הידועה לו של הלקוח בנקודות מסוימות (העיגולים הכחולים באיור). השרת (איור אמצעי) מבצע את החישוב ומעביר את הערכים חזרה אל הלקוח לבדיקה. הלקוח (איור שמאלי) מבצע התאמה של הערכים שהתקבלו מהשרת לערכי הפונקציה הסודית שברשותו. במידה והערכים קרובים מספיק לערכי הפונקציה – הלקוח יוכל להניח שהפונקציה ידועה לשרת.

איור 2: הוכחה באפס ידיעה - דוגמה למימוש

התהליך מאפשר לוודא שהשרת הוא בעל ידע או תכונה (במקרה זה – ידע אודות הפונקציה), אך מבלי שהלקוח נדרש להציג או לשלוח מידע זה לצד שלישי.

הוכחה באפס ידיעה מסייעת בהקטנת חשיפת מידע אישי, שכן בזכות מנגנון זה ניתן להשיג את המטרה המבוקשת ללא צורך בשיתוף מלוא המידע עם גורם נוסף. לצד זאת, נדרש תכנון של הפרוטוקול המתאים לכל מקרה, ובהתאם לכך שימוש בשיטה זו מתאים לשאילתות ממוקדות

ppa@justice.gov.il | 073-3928529 | *3103
קרית הממשלה, ת.ד. 33503, תל-אביב 6133401 | www.go.gov.il/ppa

והגנה על מידע בעל ערך ורגישות. מאחר שהוכחה באפס ידיעה מספקת רמת וודאות מסוימת ולא ידיעה מוחלטת באשר לנכונות המידע – חשוב להתאים את רמת הוודאות הנדרשת (ובהתאם – מספר החישובים או קושי משימת ההערכה) ליישום הספציפי.

הוכחה באפס ידיעה – מקורות להרחבה
1. בלוג "מדע גדול, בקטנה" הכולל מידע נוסף ודוגמאות בנושא הוכחה באפס ידיעה: הוכחה באפס ידיעה
2. הרצאה של בנושא הוכחה באפס ידיעה של פרופ' אלון רוזן (אוניברסיטת רייכמן): Introduction to Zero Knowledge
3. הרצאה בנושא הוכחה באפס ידיעה (טכנית) של פרופ' שפי גולדווסר (MIT, מכון וייצמן): Introduction and History of ZKP

חישוב רב-משתתפים – Multi-Party Computation

חישוב רב-משתתפים הוא פרוטוקול הצפנה המאפשר חישוב או פעולה על מידע אישי של מספר משתתפים, כך שהמידע של כל משתתף יהיה חשוף אליו בלבד. כל משתתף יוכל לקבל את תוצאת החישוב המשותף בהתאם לסוג הפרוטוקול הממומש ואופן העיבוד. דוגמה לבעיה קלאסית לחישוב רב-משתתפים היא "בעיית המיליונרים" – קבוצת אנשים המעוניינים לדעת למי ההון המקסימלי ביניהם אך מבלי לגלות כל מידע, גם לא ההון שלהם עצמם, לאף אחד מהאחרים.

דרך אפשרית למימוש חישוב רב-משתתפים היא שיתוף סוד (secret sharing). טכניקה זו מפצלת את הנתון שעליו מבקשים להגן לכמה חלקים ומפיצה את החלקים בין המשתתפים בחישוב. צד אחד לא יקבל מידע של צד אחר, אלא אם כל חתיכות המידע אשר הופצו לצדדים יאוחדו. שיטה זו מקטינה את הסבירות לכך שהמידע ייחשף במלואו, משום שיש להגיע למידע שבידי כל אחד מהצדדים על מנת לקבל גישה לכלל המידע.

דוגמה למימוש חישוב רב-משתתפים

נניח ששלושה עובדים (איילת, בן וג'סיקה) מעוניינים לחשב את ממוצע משכורתם מבלי לגלות לאחרים את המשכורת שלהם עצמם. לצורך כך, כל עובד יחלק את משכורתו לשלושה חלקים אקראיים (כך שסכום חלקים אלו יהיה שווה למשכורת שלהם). לדוגמה, אם השכר של איילת הוא 15 אש"ח, היא תוכל לחלק אותו לשלושה חלקים של 9, 2 ו-4 אש"ח. דוגמה לערכים אפשריים עבור שלושת העובדים מופיעה בטבלה הבאה:

עובד	שכר	חלק א'	חלק ב'	חלק ג'
איילת	15	9	2	4
בן	18	20	10	-12
ג'סיקה	12	8	1	3

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 קרית הממשלה, ת.ד. 33503, תל-אביב

בשלב הבא כל עובד ישמור לעצמו חלק אחד ויעביר שני חלקים נוספים למשתתפים האחרים בחישוב (למשל, איילת תעביר את חלק ב' שלה לבן וחלק ג' שלה לג'סיקה). כעת החלקים אצל המשתתפים יהיו:

עובד	חלק א' (נשמר אצל העובד)	חלק שהתקבל	חלק שהתקבל	סכום החלקים
איילת	9	10 (מבן)	1 (מג'סיקה)	20
בן	20	2 (מאיילת)	3 (מג'סיקה)	25
ג'סיקה	8	4 (מאיילת)	-12 (מבן)	0

בשלב האחרון כל עובד יחשב את הסכום של החלקים שאצלו (20, 25 ו-0 במקרה שלנו), והממוצע יתקבל מחלוקת סכומם במספר המשתתפים (3). כך שקיבלנו את התוצאה הנכונה (15), מבלי שאף אחד מהמשתתפים ראה את נתון השכר של המשתתפים האחרים או יכול להסיק אותו מהמידע שעבר אליו וממנו.

מימוש פרוטוקול חישוב רב-משתתפים יכול להיעשות במספר תצורות, בהתאם למספר המשתתפים בחישוב ורמת האמון בהם. ישנן תצורות המאפשרות להתגבר על מספר משתתפים שאינם מהימנים ואף כאלו המפעילים במכוון התקפות על האלגוריתם. כלי זה מיושם במספר רב של יישומים מעשיים ומאפשר רמת הגנה גבוהה על המידע.

חישוב רב-משתתפים דורש מאמצי חישוב ותקשורת מרובים כחלק מהתהליך של חלוקת הסוד בין המשתתפים בחישוב. הכלי יכול לכלול הצפנה הומומורפית או מנגנונים מתמטיים אחרים לביזור מידע. שימושים אפשריים של חישוב רב-משתתפים יכולים לכלול משימות מורכבות ומשמעותיות כגון הצבעה אלקטרונית או כריית מידע.

חישוב רב-משתתפים – מקורות להרחבה
1. תיאור של נושאי חישוב רב-משתתפים באתר של Institute of Electrical and Electronics Engineers - IEEE: What Is Multiparty Computation?
2. הרצאה של פרופ' טל רבין (אוניברסיטת פנסילבניה) בנושא חישוב רב-משתתפים: Secure Multiparty Computation
3. מאמר באתר Medium בנושא חישוב רב-משתתפים: A Crash Course on MPC

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 קרית הממשלה, ת.ד. 33503, תל-אביב

הצלבת מערכי מידע – Private Set Intersection

הצלבת מערכי מידע היא מקרה פרטי של חישוב רב-משתתפים ומאפשרת לשני צדדים למצוא אלמנטים משותפים בשני מערכי הנתונים, מבלי לחלוק או להעביר מידע על האלמנטים שאינם משותפים זה לזה. ניתן להשתמש בטכניקה זו כדי לחשב את גודל ההצטלבות (מספר נקודות המידע אשר תאמו זו את זו בין שני הצדדים) ולבצע חישובים סטטיסטיים על הקבוצה המשותפת.

הצלבת מערכי מידע אפשרית בשני אופנים:

1. **הצלבת מערכי מידע מסורתית** (Traditional Private Set Intersection): הצדדים מתקשרים באופן ישיר זה עם זה ולכל צד יש עותק של מערך הנתונים שלו.
2. **הצלבת מערכי מידע מואצלת** (Delegated Private Set Intersection): פעולות המחשוב או אחסון מערכי הנתונים מואצלים לצד שלישי.

דוגמה למימוש הצלבת מערכי מידע
נניח שלשני צדדים יש סדרה של שמות שברצונם להשוות: דנה: {משה, חיים, דפנה} בן: {יפית, עומר, חיים} אלגוריתם נאיבי (שאינו עושה שימוש במנגנוני הצפנה מודרניים מקובלים) להצלבת מערכי מידע ישתמש בפונקציית ערבול (Hashing Function), ישווה בין הערכים המתקבלים ויציג את הפריטים הזהים. לדוגמה הפריט המשותף (חיים) מודגש: דנה: {MHGH14, **343BCD**, 33AS12} בן: {**343BCD**, 67ERT4, Q67887} דוגמה זו משמשת להדגמה של עקרון הפעולה בלבד; יישום הצלבת מערכי מידע דורש מימוש של כלל האלגוריתמים ומנגנוני אבטחת המידע הרלוונטיים שפירוטם חורג מהיקף מסמך זה.

הצלבת מערכי מידע היא שיטה המשמשת ליישומים רבים. דוגמה אחת היא איתור אנשי קשר באפליקציות מסרים או רשתות חברתיות, כלומר איתור חברים משותפים מבלי לשתף מידע על חברים שאינם משותפים. באופן כללי, יישומי הצלבת מערכי מידע כוללים התאמה (כגון מציאת התאמה במאגרי גנום אנושי), הצלבת מאגרים (כגון רשימות אנשי קשר) או מדידת נתוני המרה לפרסום מקוון.

יישום הצלבת מערכי מידע מתבסס על ביצוע מדויק של פרוטוקולי הצפנה. בהפעלת הצלבת מערכי מידע חשוב להתייחס להיבטים של גודל היקף הנתונים, רמת האמון בצדדים המשתפים מידע ושילוב טכנולוגיות מגבירות-פרטיות נוספות.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 קרית הממשלה, ת.ד. 33503, תל-אביב

הצלבת מערכי מידע – מקורות להרחבה
1. הרצאה בנושא הצלבת מערכי מידע באתר של המכון הלאומי האמריקאי לתקנים וטכנולוגיה (National Institute of Standards and Technology – NIST): A Brief Overview of Private Set Intersection 2. ניתוח (טכני) של יישומי הצלבת מערכי מידע: What are we PSIgning up for? Analyzing Applications of Two-Party Private Set Intersection

למידה מבוזרת – Federated Learning

למידה מבוזרת מאפשרת לכמה גורמים במשותף לאמן מודלים של בינה מלאכותית על מידע שלהם (מודל מקומי), ולאחר מכן לשלב את הדפוסים שזוהו במודלים המקומיים אל תוך מודל גלובלי יחיד ומדויק, ללא צורך לחלוק את המידע אשר שימש כל אחד מהגורמים לאימון המודל.

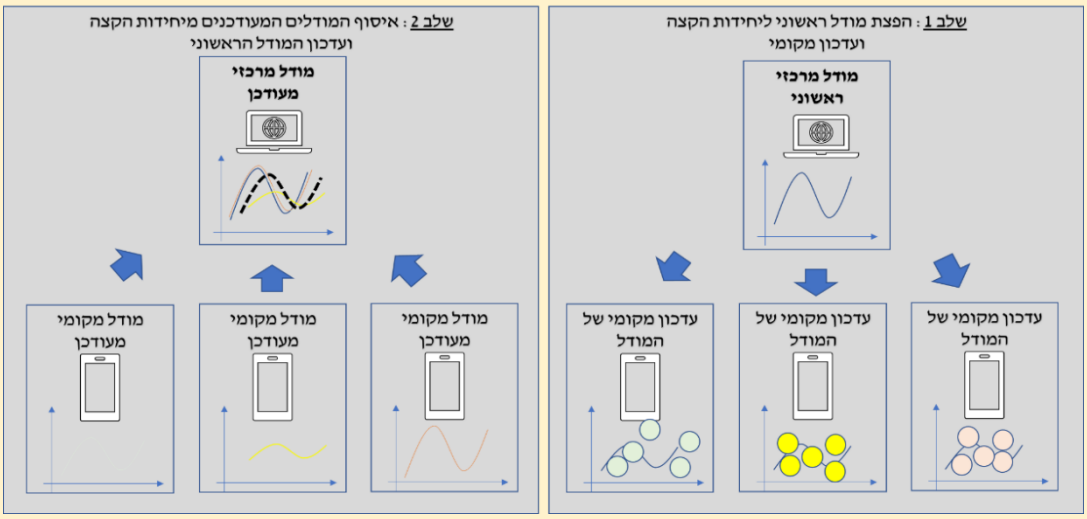
למידה מבוזרת יכולה להיעשות בגישות הבאות:

1. **למידה מבוזרת מרכזית** (Centralized federated learning): שרת מרכזי מייצר אלגוריתם או מודל, ושולח את המודל למקורות מידע מבוזרים. המודל מתעדכן בהתאם למקורות המידע המקומיים ונשלח חזרה לשרת המרכזי שיוצר מודל משוקלל.
2. **למידה מבוזרת מקומית** (Decentralized federated learning): במודל זה לא מעורב שרת מרכזי. הצדדים מתקשרים זה עם זה באופן ישיר ומעדכנים את המודל בכל פעם על סמך המידע המקומי שלהם.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 קרית הממשלה, ת.ד. 33503, תל-אביב

למידה מבוזרת מרכזית – תהליך המימוש

חברות רבות עושות שימוש בלמידה מבוזרת לאימון מערכות בינה מלאכותית תוך שמירה על המידע האישי והפרטיות של המשתמשים. תהליך נפוץ של למידה מבוזרת מרכזית מוצג באיור 3:



איור 3: שלבים בלמידה מבוזרת

השלב הראשון באיור 3 (מימין) הוא בניה של מודל ראשוני והפצה ליחידות הקצה. יחידות הקצה מאמנות ומשלימות את המודל על בסיס המידע שברשותן. בשלב השני באיור 3 (משמאל) יחידות הקצה שולחות את המודל המעודכן ליחידה המרכזית. היחידה המרכזית משקללת את המודלים ומייצרת מודל מעודכן. התהליך יכול לחזור על עצמו מספר פעמים.

השימוש בלמידה מבוזרת יוצר עומס חישובי שיכול להיות משמעותי בעיבוד מידע בהיקפים רחבים. יישום למידה מבוזרת נדרש להתייחס למאפייני מאגר המידע שבשימוש, סיכוני זיהוי חוזר מתוך המודלים והיבטי שילוב טכנולוגיות מגבירות-פרטיות נוספות להגברת ההגנה על המידע האישי.

למידה מבוזרת – מקורות להרחבה
1. הרצאה של ד"ר אביב קרן בסדנת יובל נאמן למדע, טכנולוגיה וביטחון באוניברסיטת תל אביב: Federated Learning in the real world 2. בלוג באתר OpenMined בנושא למידה מבוזרת: Design a federated learning system in seven steps

ppa@justice.gov.il | 073-3928529 | *3103
קרית הממשלה, ת.ד. 33503, תל-אביב 6133401 | www.go.gov.il/ppa

סביבת ביצוע מהימנה – Trusted Execution Environment

סביבת ביצוע מהימנה היא אזור מאובטח במחשב, בשרת או בטלפון הנייד, לעיבוד מידע רגיש. סביבת ביצוע מהימנה מאפשרת להריץ קוד ולגשת למידע באופן נפרד ומבודד מיתר חלקי המערכת. המימוש של סביבת ביצוע מהימנה מבוסס על הפרדה פיזית או לוגית מיתר חלקי המערכת, כך שלתוכנות ומשתמשים מחוץ לסביבה זו אין גישה למידע המעובד בה. סביבת ביצוע מהימנה מבטיחה כי ישויות לא מורשות לא תוכלנה לראות, לשנות, לעדכן או להסיר את המידע או הקוד בשימוש בסביבה זו. טכנולוגיה זו מאפשרת להגן מפני גישה של גורמים בלתי מורשים למידע בעודו בשימוש בתוך הסביבה.

דוגמה: סביבת ביצוע מהימנה – תשתית ושימושים

מערכות תעשייתיות רבות עושות שימוש בסביבת ביצוע מהימנה. התשתית לסביבת ביצוע מהימנה מוטמעת במעבדים של חברות כגון אינטל, AMD, NVIDIA ואחרות. השימוש בסביבות ביצוע מהימנות נפוץ גם במכשירים ניידים ובמערכות הפעלה מרכזיות, כגון iOS ואנדרואיד, הכוללות תמיכה בטכנולוגיה זו.

הפיתוח של סביבות ביצוע מהימנות החל ממנגנונים מבוססי חומרה, שבהם מוטמעות יכולות ומפתחות הצפנה ייעודיים. בשנים האחרונות מפותחות מערכות ומנגנוני זיהוי, אימות והצפנה המאפשרים עבודה עם סביבת ביצוע מהימנה המבוססת תוכנה.

שימושים נפוצים בסביבת ביצוע מהימנה כוללים:

1. **זיהוי והזדהות ביומטרית**: התאמה של מידע ביומטרי מול מנוע זהויות כחלק מתהליכי וידוא זהות ביומטריים (זיהוי פנים, טביעות אצבע וקול). סביבת ביצוע מהימנה מספקת הגנה על המידע הביומטרי ומייצרת חוצץ כנגד אפליקציות לא מאובטחות הפועלות במערכת.
2. **חישוב מאובטח**: ביצוע חישוב ללא מתן יכולת לגשת למידע או קוד המשמשים לצורך ביצוע החישוב. סביבה מאובטחת לביצוע קוד רגיש, כגון אלגוריתמים להצפנה או תוכנות רגישות למידע המבטיחים שהקוד הרגיש לא ייחשף או ישונה על ידי תוכנות אחרות או תוקפים.
3. **שירותים פיננסיים**: מתן גישה מאובטחת לארנקים ופרטי אשראי לצורך תשלום, בפרט במכשירים ניידים. ארנקים דיגיטליים, כמו Google Pay או Apple Pay, משתמשים בסביבת ביצוע מהימנה כדי לאחסן ולהגן על מפתחות הצפנה הנדרשים לביצוע תשלומים בצורה מאובטחת.
4. **ניתוח מידע ולמידת מכונה**: ביצוע חישובים באופן המגן על הפרטיות, בכלל זה חישוב רב-משתתפים – Multi-Party Computation ולמידה מבוזרת – Federated Learning.

ppa@justice.gov.il | 073-3928529 | *3103
קרית הממשלה, ת.ד. 33503, תל-אביב 6133401 | www.go.gov.il/ppa

סביבת ביצוע מהימנה – מקורות להרחבה
1. מסמך של The Confidential Computing Consortium בנושא סביבת ביצוע מהימנה: Confidential Computing: Hardware-Based Trusted Execution for Applications and Data 2. פרויקט של הלשכה האירופית לסטטיסטיקה (Eurostat) לשימוש בסביבת ביצוע מהימנה לביצוע עיבודים סטטיסטיים: Project ESTAT.2019.0232

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 קרית הממשלה, ת.ד. 33503, תל-אביב

ניתן לשפר את רמת ההגנה על מידע אישי הנמצא במאגר מידע באמצעות מנגנונים שיש בהם כדי לאפשר לנושא המידע שליטה טובה יותר בהרשאות הגישה למידע האישי ויכולת מעקב נוחה אחר הגישות בפועל למידע האישי אודותיו.

מאגרי מידע בשליטת נושא המידע – Personal Data Stores

מאגרי מידע בשליטת נושאי המידע הם אמצעים המשלבים מנגנונים למעקב אחר הגישה למידע האישי. אלו הם פתרונות טכנולוגיים שמאפשרים למשתמשים לנהל, לשמור ולהשתמש בנתונים האישיים שלהם בצורה מאורגנת ומאובטחת. מאגרים אלו נועדו להחזיק מידע עבור אפליקציות ושירותים שונים, ומאפשרים למשתמשים גישה ופיקוח על השימוש הנעשה במידע שלהם.
מאגרי מידע אלו עשויים לכלול מידע ממקורות שונים, כמו פרטי קשר, היסטוריה רפואית, פרטי חשבון או מסמכים. רבים מהמאגרים פועלים על גבי פלטפורמות ענן ובהם הנתונים מוצפנים ומוגנים כדי להבטיח גישה רק למשתמשים המורשים. אפשרות נוספת היא אחסון מקומי במכשירים האישיים של המשתמשים, כגון מחשבים או טלפונים ניידים, עם אמצעים לאבטחת המידע באופן מקומי.
משתמשים יכולים לשלוט ולנהל את הרשאות הגישה למידע שלהם. הדבר מאפשר למשתמשים לשתף נתונים עם אחרים בצורה מאובטחת אם הם בוחרים לעשות זאת. במקרים מסוימים, מאגרים אלו מאפשרים שיתוף מידע עם שירותים או אפליקציות אחרות בצורה מבוקרת, כאשר המשתמש נותן לכך אישור מפורש. משתמשים יכולים למחוק או לעדכן את המידע שלהם בהתאם לצורך ולשמירה על פרטיותם.

דוגמה: מאגרי מידע בשליטת נושאי המידע – יישום אפשרי

באופן מסורתי מידע פיננסי נשמר בארגון הרלוונטי להפקתו או השימוש בו – בנק, מקום עבודה (משכורת), חברת כרטיסי אשראי. בצורה זו העברת המידע בין מוסדות שונים היא חלקית ונדרשת לביצוע בכל פעם מול מוסד אחר הנדרש למידע.

שימוש במאגרי מידע בשליטת נושאי המידע עבור מידע פיננסי עשוי לאפשר לאדם לשמור את המידע במקום אחד תוך מתן הרשאה לגוף הנדרש לגשת למידע בכל פעם. דוגמה לשימוש אפשרי הוא מתן הרשאת צפיה בנתוני שכר ותנועות עובר ושב לבנקי משכנתאות לצורך בחינת זכאות להלוואה. כך העברת המידע תהיה פשוטה ולאדם תהיה היכולת לשלוט ולעקוב אחר הרשאות הגישה.

מאגרי מידע בשליטת נושאי המידע הם פתרונות מתקדמים שמספקים למשתמשים שליטה על המידע האישי שלהם, ומאפשרים להם לנהל אותו ואת הרשאות הגישה אליו בצורה מאורגנת

ppa@justice.gov.il | 073-3928529 | *3103
קרית הממשלה, ת.ד. 33503, תל-אביב 6133401 | www.go.gov.il/ppa

ומאובטחת. לצד הנוחות בשימוש במאגרים אלו, נדרשת מצד המשתמשים רמת אחריות משמעותית בניהול ובקרה אחר הרשאות הגישה, והקפדה על אימות המורשים לגשת למידע האישי.

מאגרי מידע בשליטת נושאי המידע – מקורות להרחבה
1. סקירה בנושא מאגרי מידע בשליטת נושאי המידע של משרד הבריאות האמריקאי: Personal Data Stores (PDS): A Review 2. דיווח של ה-BBC על מערכת מאגרי מידע שמנהלת פרופיל מדיה אחוד עבור הצופה: Personal data stores: building and trialling trusted data services 3. מאמר באתר Medium בנושא מאגרי מידע בשליטת נושאי המידע: What IS a Personal Data Store?

כלי תיעוד ושקיפות – Documentation and Transparency Tools

ניתן לשפר את רמת ההגנה על מידע אישי הנמצא במאגר מידע באמצעות מנגנונים שיש בהם כדי להבטיח תיעוד מדויק ומלא של הגישה למידע (באופן שאינו ניתן לשינוי), תוך מתן אמצעים יעילים לנושא המידע למעקב אחר השימוש הנעשה במידע האישי אודותיו. שילוב אמצעים אלו מסייע ביצירת אמון ותרבות של שמירה על הגנת הפרטיות מעצם תיעוד הפעולות ומתן הגישה של נושא המידע למידע על הפעולות המבוצעות במידע האישי אודותיו.

כלי תיעוד ושקיפות הם אמצעים מתפתחים המיועדים להגביר את הפרטיות על ידי מתן דרכים חדשות לשליטה טובה יותר של נושאי המידע על המידע האישי שלהם.

כלי תיעוד ושקיפות – דוגמה ליישום – פורטל Data Tracker
ממשלת אסטוניה מפעילה פורטל מעקב אחר השימוש הנעשה במידע האישי (Data Tracker Portal)[20] המציג מידע על גישות למידע אישי המוחזק במאגרי מידע ממשלתיים. הפורטל מציג תיעוד של הגישות שבוצעו, כולל תאריך ושעה של הגישה למידע, פרטי הארגון מבקש הגישה, מאגר המידע שאליו בוצעה הגישה וסוג השאילתה או התהליך שעבורו נדרש המידע). דוגמאות לשימוש לא נאות במידע שדווחו על ידי האסטונים באמצעות שימוש בתשתית הדיגיטלית המאפשרת לאזרח לצפות בפרטי הגישה למידע האישי שלהם כוללים אירועים של שוטר שניגש למידע אישי אודות אשתו לעתיד וחובש שבדק מדוע הוזעק אמבולנס לכתובת ספציפית לבקשת שכנה סקרנית[21].

[20] פירוט על פורטל מעקב המידע האסטוני באתר של e-Estonia: Data tracker – tool that builds trust in institutions.

[21] תיאור בכתבה באתר של e-Estonia: https://e-estonia.com/i-spy-with-my-little-eyeprivacy/.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 קרית הממשלה, ת.ד. 33503, תל-אביב

דוגמה נוספת לכלי תיעוד ושקיפות מידע הוא מרשם המידע הרפואי באוסטרליה. במערכת זו מתאפשר לאזרח לקבל הודעה כאשר מתבצעת גישה לרשומה שלו או כאשר מתבצעים סוגים מסוימים של שינויים. מאחר וההתרעה מופקת בעת הגישה, לארגון שביצע זאת אין אפשרות להכחיש מאוחר יותר את עצם הגישה למידע האישי. מידע נוסף על המערכת מופיע בקישור.

סיכום

השימוש בטכנולוגיות מגבירות-פרטיות מסייע בהגנה על הפרטיות במגוון שיטות ואמצעים. שילוב טכנולוגיות להגנה על הפרטיות במוצרים ומערכות נדרש להיעשות כחלק מראיה מערכתית של המידע שבשימוש, רמת הסיכון לפרטיות, התהליכים, הרגישות והכלים הארגוניים והמשפטיים המשלימים. הרחבת השימוש בטכנולוגיות אלו תסייע בשיפור ההגנה על הפרטיות כחלק ממאמץ כולל של הגנה על המידע בעידן הדיגיטלי ולצד האמצעים המשפטיים המקובלים לאכיפת דיני הגנת הפרטיות.

ppa@justice.gov.il | 073-3928529 | *3103
קרית הממשלה, ת.ד. 33503, תל-אביב 6133401 | www.go.gov.il/ppa