ז' סיון, תשפ"ה

3 ביוני 2025

טיוטת הנחייה על תחולת הוראות החוק על מערכות AI – הערות FPF ישראל

עו"ד רבקי דב"ש, עמיתה בכירה, FPF-Israel (המכון הישראלי למדיניות טכנולוגיה)

טיוטת ההנחייה של הרשות להגנת הפרטיות בנושא "תחולת הוראות חוק הגנת הפרטיות על מערכות בינה מלאכותית" (להלן – הטיוטה) עוסקת באופן שבו חוק הגנת הפרטיות הקיים ותקנותיו, חלים על מערכות בינה מלאכותית. הטיוטה מתייחסת, בין היתר, לעקרונות הבאים:

1. בכל שלב של חיי מחזור החיים של מערכת בינה מלאכותית, ובכלל זה אימון המודל, נדרש שיהיה בסיס חוקי לעיבוד המידע האישי.
2. לשם עמידה בדרישת החוק ל"הסכמה" מדעת ביחס לשימוש במערכת הבינה המלאכותית, נדרש בעל השליטה, לשיטת הרשות –
   2.1 לכלול תיאור על אופן פעילות המערכת ברמת הפירוט הנדרשת ובשים לב למגבלות טכנולוגיות;
   2.2 להבהיר האם האינטראקציה של נושא המידע היא מול מערכת בינה מלאכותית, ככל שזו עובדה רלוונטית להסכמה;
   2.3 פירוט פרטי וסוגי המידע בהן עושה המערכת שימוש, ומקורותיהם;
   2.4 ההסבר נדרש ביחס לכל אחת ממטרות השימוש, כולל אימון האלגוריתם, ופירוט הסיכונים הכרוכים בשימוש במערכת;
3. כריית מידע מרשת האינטרנט לצורך אימון המערכת אסור ללא הסכמה מדעת של נושא המידע. אין להניח כי מי שפרסם את המידע אודותיו ברשת הסכים מכללא גם לשימוש במידע אודותיו לאימון מערכות בינה מלאכותית. על בעלי אתרים מסוג זה להטמיע מנגנונים שיחסמו את אפשרות כריית המידע מאתרים אלו.
4. על מנת לפעול באופן אחריותי, ממליצה הרשות לגופים העושים שימוש במערכות בינה מלאכותית למנות ממונה הגנת הפרטיות (DPO) ולערוך תסקיר השפעה על הפרטיות. כמו כן הרשות מצפה כי הדירקטוריונים באותם ארגונים יפעלו לקבוע את מדיניות השימוש של הרשות במערכות אלו.
5. ביחס לזכויות נושא המידע מבהירה הרשות כי לשיטתה –
   5.1 הזכות לתיקון מידע שגוי מכח החוק כוללת גם את הזכות לבקש את תיקון האלגוריתם שהוביל למידע השגוי;
   5.2 ביחס למאגרים הכוללים מידע מהאזור הכלכלי האירופאי, והנדרשים לפעול באופן אקטיבי למניעת מידע שגוי, עליהם לייצר מנגנונים המצמצמים את הסיכוי כי המערכת תלמד או תפעל באופן שגוי.

נייד: 6217710 50 972 +

www.techpolicy.org.il| rdvash@techpolicy.org.il

6. במימוש חובות אבטחת המידע, על בעלי השליטה להתייחס למאפיינים הייחודיים בתחום זה למערכות בינה מלאכותית.
7. ככל שהמאגר חייב ברישום, יש לפרט בעת הרישום את מטרות השימוש בבינה המלאכותית.

במהלך חודש נובמבר 2024, כינס ה-FPF, ארגון האם של המכון הישראלי למדיניות טכנולוגיה, שולחן עם מגוון רחב של מומחים טכנולוגיים ורגולטורים ממדינות שונות, בניסיון ללמוד במשותף על ההיבטים הטכניים של מערכות LLM, וביצד ניתן להבטיח עיבוד חוקי של נתונים אישיים ועמידה בחובות מכח רגולציית הגנת המידע האישי במערכות אלו.

אנו סבורים כי הסיכום של השולחן העגול שנערך ב-FPF, יכול לספק תובנות טכניות חשובות שעשויות להעשיר את טיוטת ההנחייה בנקודות מסוימות, כמפורט להלן.

1. **מגבלות טכניקות להגנת הפרטיות**

הטיוטה מתייחסת לצורך בצמצום מידע (סעיף 9.6), כמו גם בצורך להדגיש את סיכוני השימוש במערכות בינה מלאכותית אגב חובת השקיפות והיידוע (סעיף 6.2).

מהדיון בשולחן העגול עלה כי טכניקות שונות המיועדות לצמצום סיכוני פרטיות המוכרות לנו מעולם בסיסי הנתונים, כמו פרטיות דיפרנציאלית, נתונים סינתטיים או דה-דופליקציה (deduplication),[1] עלולות לעמוד בפני מגבלות בהקשרים של מערכות AI. לדוגמה, פרטיות דיפרנציאלית עשויה לפגוע בביצועי המודל, במיוחד ביחס למידע נדיר אך בעל ערך, ונתונים סינתטיים אינם מבטיחים פרטיות לבדם ואף יכולים להוביל להתנהגויות בלתי צפויות של המודל. כמו כן, דה-דופליקציה אינה תמיד קלה ליישום בשל הקושי להגדיר כפילויות בריבוי מקורות מידע מגוונים.

לפיכך מוצע שהטיוטה תתייחס לכך שככל שנעשה שימוש בטכניקות להגברת ההגנה על הפרטיות, תדרש הערכה בדבר ההשלכה שיש לטכניקות אלו ובחינה האם הפגיעה במהימנות המודל (שהיא עצמה עשויה לפגוע בפרטיות, לתפיסת הרשות, ככל שתוביל להסקת מסקנות לא מהימנות) - מוצדקת. הערכה זו מוצע שתבוצע במסגרת תסקיר הפרטיות, או מסמך המאגר. מכל מקום, יש לתת גילוי דעת על הפשרות שנעשו במודל, ככל שנעשו, במסגרת היידוע של נושא המידע. נציין כי מחקר יישומי משמעותי באקדמיה ובתעשייה מתמקד כיום במאמצים יעילים יותר ליישום טכניקות לשיפור הפרטיות על נתוני אימון, וההתקדמות בתחום זה מצדיקה תשומת לב רבה להשלכות על ההערכות.

1 בהקשר למערכות AI הכוונה היא להסרה של מופעים כפולים או חוזרים על עצמם של נתונים מתוך מאגרי המידע שמשמשים לאימון המודל.

2. מקור נתוני אימון – תמריץ לשימוש בנתוני משתמשים

הטיוטה מתייחסת למגבלות על שימוש במידע אישי המצוי ברשת. עמדה זו תואמת את המדיניות ברבות ממדינות אירופאיות (כפי שאף מוער בה"ש בטיוטה).

בשולחן העגול הוער כי מחסור בנתונים פתוחים באיכות גבוהה דוחף חברות לשימוש מוגבר בנתוני משתמשים לאימון באופן שמאתגר את המוסכמות לשימוש בנתוני משתמשים בהתאם לדיני הפרטיות המקובלים.

לפיכך מוצע כי הרשות תתן דעתה להנחיית בעלי השליטה המבקשים לעשות שימוש במערכות AI, תוך "אימון" המערכת על בסיס הנתונים של נושאי המידע במאגר שבידי בעל השליטה. מוצע להבהיר מהי לתפיסת הרשות מסגרת השימוש הלגיטימית, ומה הם מגבלות השימוש במאגר שבידי בעל השליטה.

3. הדגשת השלכות הפרטיות של שלבי אימון ספציפיים

הטיוטה מתייחסת לצורך בעמידה בהוראות החוק בכל אחד מ"שלבי מחזור החיים של המערכת" (סעיף 5.3).

בשולחן העגול ישנו פירוט לגבי שלבי האימון השונים של המערכת, תוך התייחסות לשלב של ה"fine-training" כבזה המציג סיכוני פרטיות גבוהים במיוחד בשל רגישות הנתונים שנדרשים לשם דיוק המערכת.

לפיכך מוצע כי הנחיית הרשות תעמיק את הפירוט של שלביה השונים של מערכת AI, תוך הדגשת הסיכונים הייחודיים לכל שלב, וקיומן של דרישות הגנה מותאמות יותר לכל שלב של המערכת - לדוגמה בהקשר של דרישות אבטחת מידע או הצורך בתסקיר השפעה על פרטיות.

4. האתגר הטכני של מימוש הזכות לתיקון מידע

הטיוטה מתייחסת לזכות נושא המידע לבקש תיקון או מחיקה של מידע שגוי, וקובעת כי זכות זו עשויה להתייחס גם לתיקון האלגוריתם של המערכת (פסקה 8).

בדיון בשולחן העגול עלה הקושי של "מחיקת" מידע ממודל LLM לאחר אימונו ("unlearning") , וזאת בשונה מיישום הדרישה למחיקת מידע מבסיסי נתונים. שיטות "אי-למידה" קיימות מתייחסות למנעד פתרונות: מרשת אימון מחדש (יקרה מאוד) ועד גישות יעילות יותר אך מורכבות או

תיאורטיות יותר כמו Sharding או "אי-למידה מקורבת" (approximate unlearning).[2] המורכבות המתוארת בסיכום השולחן העגול אינה עולה לידי ביטוי בטיוטת ההנחייה, המתייחסת ליישום של הזכות לתיקון מידע והסרת מידע שגוי בעניין פשוט.

לפיכך מוצע שהנחיית הרשות תכלול התייחסות מפורשת לאתגרים הטכניים והעלויות המשמעותיות הכרוכות במימוש זכות התיקון. מוצע להבהיר בהנחייה כיצד הרשות מצפה שבעלי השליטה יעמדו בדרישות אלה, תוך הכרה בכך שייתכן שיידרשו גישות טכניות מתקדמות שאינן תמיד טריוויאליות ליישום. ייתכן שיהיה מקום לשקול (בהתאם לרמת הסיכון וליכולת הטכנולוגית המוכחת) האם קיימות נסיבות בהן גישות של "אי-למידה מקורבת" יכולה להיחשב כעמידה סבירה בדרישה, בכפוף למתן הסברים ושקיפות לנושאי המידע והרשות.

5. עידוד פיתוח כלים טכניים למימוש זכויות נושאי המידע

טיוטת ההנחייה מתייחסת גם לזכויות נושא המידע (פסקה 8).

במהלך השולחן העגול, נדון הצורך בתפיסה הממוקדת יותר במשתמש, ובכלל זה עידוד פיתוח כלים טכניים ו-API שיאפשרו לנושאי המידע לשלוח שאילתות למודלים לגבי מידע אישי, ליצור דוחות פרטיות אוטומטיים ומנגנוני התאמה.

לאור המורכבות של מערכות AI והצורך ליידע את נושא המידע, מוצע לרשות לשקול התייחסות לחשיבות של פיתוח כלים טכניים שיקלו על נושאי המידע לממש את זכויותיהם באופן יעיל. בתחומים שבהם מימוש זכויות כמו גישה או מחיקה אינו מעשי כיום, יש לתת את הדעת לחלופות שמשיגות את מטרות ההגנה על הפרט, כגון על ידי דיכוי תוצאות (מניעת יציאת מידע הקשור לאותו אדם מתוך המודל, אף על פי שהמידע עדיין קיים בו מבחינה טכנית).

6. לסיכום

השימוש במערכות AI הולך וגובר. ישנה חשיבות בבחינת ההוראות החלות על מערכות אלו גם ביחס לדיני הגנת המידע האישי. במסמך זה ביקשנו להציף את האתגרים שבהכוונת השימוש במערכות AI, תוך התבוננות במאפיינים טכנולוגיים ייחודיים של מערכות אלו, על בסיס דיון שנערך בארגון אליו המכון שייך.

[2] המונח "approximate unlearning"הוא מושג טכני מעולם הבינה המלאכותית שפירושו ניסיון לגרום למודל AI "לשכוח" מידע ספציפי ששימש לאימונו, על ידי שינוי הפרמטרים הפנימיים שלו באופן שמחקה את התוצאה של אימון מחדש מלא, אך בעלות או מורכבות נמוכה יותר.

מטבע הדברים, הסוגיות נכתבו בתמצית, ונשמח להרחיב ולקשר לגורמים רלוונטיים בתחום, ככל שנדרש.

נייד: 6217710 50 972 +
www.techpolicy.org.il| rdvash@techpolicy.org.il