עקרונות מנחים לפיתוח טכנולוגיות מבוססות למידת מכונה

הקדמה

לטכנולוגיות בינה מלאכותית (Artificial Intelligence, AI) ולמידת מכונה (Machine Learning, ML) יש פוטנציאל ליעל ולשפר את שירותי הבריאות על-ידי הפקת תובנות חדשות וחשובות מהכמות הגדולה של נתוני בריאות הנאספים ומצטברים בארץ ובעולם. טכנולוגיות אלו עושות שימוש באלגוריתמים כדי ללמוד ממידע בעולם האמיתי, ובמצבים מסוימים עשויות לעשות שימוש במידע זה כדי לשפר את ביצועי המוצר לאורך זמן. עם זאת, ישנם היבטים הדורשים התייחסות ייחודית בפיתוח טכנולוגיות אלו בשל המורכבות שלהן ובשל האופי המשתנה והמתפתח שלהן.

באחרונה פרסמו רשויות רגולטוריות מובילות בעולם: FDA, Health Canada וה-MHRA, עשרה עקרונות מנחים להתנהלות טובה בתחום למידת מכונה - Good Machine Learning Practice (GMLP):
https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles

עקרונות מנחים אלה יסייעו לקדם פיתוח ושימוש בטוח, יעיל ואיכותי במכשירים רפואיים אשר עושים שימוש בבינה מלאכותית ולמידת מכונה (AI/ML). ניתן דגש מיוחד לעקרונות ייחודיים ללמידת מכונה, כגון: מניעת הטיות (bias), מניעת התאמת יתר (Overfitting), שקיפות, מהימנות הנתונים, פרטיות ובטחון מידע.
פרסום זה מצטרף לשורה של מסמכי מדיניות ופעילויות בזירה הבינלאומית בתחום של בינה מלאכותית, הכוללות התייחסות ואסדרה לתחום טכנולוגי זה. בישראל, התייחסות זו מרוכזת ע"י משרד המדע והטכנולוגיה, אשר פירסם באחרונה עקרונות מדיניות ממשלתיים שינחו את דרך השימוש בטכנולוגיות הבינה המלאכותית. העקרונות המובאים במסמך זה נמצאים בהלימה לעקרונות הרוחביים המוצעים במסמך של משרד המדע והטכנולוגיה. למידע נוסף:
https://www.gov.il/he/departments/news/most-news20223110

עשרת העקרונות המנחים להלן נועדו להניח את היסודות להתנהלות טובה בתחום למידת מכונה המתייחסים לאופי הייחודי של מוצרים אלה. הם גם יסייעו לטפח צמיחה עתידית אחראית בתחום זה המתקדם במהירות. עקרונות אלו מהווים חלק מתוך תוכנית לפיתוח מתווה רחב יותר של קווים מנחים והגדרות לאמצעים רפואיים דיגיטליים (אר"ד) הנותנים מענה לייחודיות של טכנולוגיות אר"ד, תוך דיאלוג עם כלל בעלי העניין והסתמכות על הגדרות והנחיות בינ"ל. למידע נוסף:
https://www.gov.il/he/departments/units/health-innovation-committee

עקרונות מנחים להתנהלות טובה בתחום לימוד מכונה	
שילוב מומחיות רב-תחומית לאורך כל מחזור החיים של המוצר	בניית מודל המותאם לנתונים הזמינים ומשקף את השימוש המיועד במוצר
שימוש בשיטות מיטביות של הנדסת תוכנה ואבטחת מידע בבניית המודל	התמקדות בביצועים של צוות האדם-בינה מלאכותית
ייצוג ראוי של אוכלוסיית המטופלים המיועדת במסגרת משתתפי המחקר הקליני ומערכי הנתונים	בדיקת ביצועי המכשיר בתנאים רלוונטיים מבחינה קלינית
שימוש בבסיסי נתונים בלתי תלויים לאימון ולבחינה של המודל	העברת מידע חיוני למשתמשים בצורה שקופה וברורה
שימוש בשיטות מקובלות ומיטביות לפיתוח מערך נתוני ייחוס	ניטור ביצועי המודלים המוטמעים וניהול הסיכונים הכרוכים באימון מחדש

משרד הבריאות הישראלי מפרסם מסמך זה עבור גורמים העוסקים בפיתוח טכנולוגיות המבוססות על למידת מכונה (AI/ML) בתחום הבריאות במטרה לסייע בתהליכי הפיתוח, וזאת על-ידי התאמת שיטות עבודה טובות שהוכחו במגזרים אחרים והתאמתן כך שיהיו ישימות לטכנולוגיה רפואית ולמגזר הבריאות בישראל; עקרונות אלו יהוו בסיס לבנייה של שיטות חדשות ספציפיות לעולם הבריאות, אשר ימשיכו ויתעדכנו ככל שתחום המכשור הרפואי מבוסס למידת מכונה יתפתח. עבודה זו מתבצעת בתיאום, ותוך כדי דיאלוג מתמשך עם הרשויות הרגולטוריות המובילות בעולם.

אימוץ הקווים המנחים אשר פותחו ע״י הרשויות המובילות בעולם והתאמתם למערכת הבריאות הישראלית מסייע בתהליך הרמוניזציה של המסגרת הרגולטורית לטכנולוגיות מבוססות למידת מכונה, ומאפשר לטכנולוגיות אלו ניוד גלובלי אחוד , בטוח ומהיר , לטובת שיפור שירותי הבריאות לאזרחי ישראל.

עשרת העקרונות המנחים:

1. **שילוב מומחיות רב-תחומית לאורך כל מחזור החיים של המוצר** לטובת בחינה כי מכשירים רפואיים מבוססי בינה מלאכותית בטוחים, יעילים ועונים על צרכים קליניים משמעותיים לאורך מחזור החיים של המוצר. לטובת בחינה זו נדרשת הבנה מעמיקה של האופן בו עתיד המודל להשתלב בתהליך העבודה הקליני, לצד היתרונות הרצויים והסיכונים הנלווים למטופלים. גיוון אנושי בתהליך הפיתוח יכול לתת ביטוי למגוון דיסציפלינות מקצועיות, תפיסות חברתיות, וקבוצות באוכלוסייה.

2. **שימוש בשיטות מיטביות של הנדסת תוכנה ואבטחת מידע בבניית המודל** תוך מתן תשומת לב לעקרונות הבאים: שיטות הנדסת תוכנה טובות, אבטחת איכות נתונים, ניהול ותיעוד נתונים, הגנה על פרטיות, ושיטות אבטחת סייבר מעודכנות. שיטות עבודה אלה כוללות תהליך תכנון וניהול סיכונים שיטתי שיאפשר לזהות ולתקשר כראוי החלטות ורציונל של תכנון, הטמעה וניהול סיכונים, כמו גם להבטיח את מהימנות ותקינות הנתונים.

3. **ייצוג ראוי של אוכלוסיית היעד במסגרת המחקר הקליני ומערכי הנתונים** כך שניתן יהיה להכליל את תוצאות המחקר באופן סביר לאוכלוסייה הרלוונטית. בתוך כך, על פרוטוקול איסוף הנתונים להבטיח שמאפייני אוכלוסיית היעד (לדוגמה: גיל, מין, מגדר, מאפיינים סוציו-דמוגרפיים), אופן השימוש, ומדדי המודל, מיוצגים באופן ובהיקף מספקים במחקר הקליני ובמערכי הנתונים.
עקרון זה הינו חשוב כדי לנהל הטיות אפשריות, לקדם ביצועים הולמים הניתנים להכללה על פני אוכלוסיית היעד, להעריך את השימושיות ולזהות נסיבות שבהן המודל עשוי ליצר ביצועי חסר.

4. **שימוש בבסיסי נתונים בלתי תלויים לאימון ולבחינה של המודל** על-מנת להבטיח את עצמאותם. יש לקחת בחשבון ולתת מענה לכל מקורות התלות הפוטנציאליים בתהליכי הבחירה והתחזוקה של בסיסי הנתונים, לרבות: מאפייני המטופלים, רכישת נתונים, ומאפייני אתר ההרצה .

5. **שימוש בשיטות מקובלות ומיטביות לפיתוח מערך נתוני ייחוס**[1], באופן שיבטיח איסוף נתונים רלבנטיים ומאופיינים היטב מבחינה קלינית, והבנת מגבלות נתוני הייחוס. ככל שניתן, יעשה שימוש במערכי נתוני ייחוס מקובלים בפיתוח ובדיקת מודלים, המקדמים ומדגימים את יכולת ההדירות (robustness) וההכללה (generalizability) של המודל בקרב אוכלוסיית היעד .

6. **בניית מודל המותאם לנתונים הזמינים העדכניים ומשקף את השימוש המיועד במוצר** תוך תמיכה בהפחתה פעילה של סיכונים ידועים, כגון: התאמת יתר, (overfitting) ירידה בביצועים וסיכוני אבטחה. היתרונות הקליניים והסיכונים הקשורים למוצר יהיו מובנים היטב, ישמשו להפקת יעדי ביצוע קליניים משמעותיים לבחינה, ויתמכו בכך שהמוצר יוכל להשיג את השימוש המיועד בו בבטחה וביעילות. השיקולים יכללו את השונות בין הנתונים המוזנים למודל, הנתונים המופקים מהמודל, אוכלוסיית היעד ותנאי השימוש הקליניים, והשפעתם של אלו במסגרת ביצועי המודל בזירה המקומית ובזירה הבינלאומית. יש מקום להתייחס למימד הזמן של הנתונים הזמינים, מתוך הבנה כי נתונים אלו משקפים ניסיון קליני מתפתח ומשתכלל על-פני זמן.

---

[1] reference dataset - קבוצות לוגיות של ערכי קוד המהווים סטנדרט הנגיש למשתמשים שונים באופן מרוכז, כמו לדוגמה, קודי אבחנות

7. **התמקדות בביצועים של צוות האדם-בינה מלאכותית** כאשר למודל יש "גורם אנושי בתהליך" (human-in-the-loop), יש לתת דגש לביצוע של צוות האדם-בינה מלאכותית, תוך התייחסות לגורם השיקול והפרשנות האנושית. במסגרת ביצועים אלו יש מקום להתייחס לממשק העבודה של האדם עם המודל, וליכולת הפרשנות של האדם את המודל (interpretability).

8. **בדיקת ביצועי המכשיר בתנאים רלוונטיים מבחינה קלינית** ללא תלות במערך נתוני האימון. פיתוח והוצאה לפועל של תכניות בדיקה מבוססות סטטיסטית בכדי להסיק כי המכשיר משיג את התוצאות הקליניות הרצויות בסביבה הקלינית הרלבנטית. השיקולים בבניית תוכנית הבדיקה יכללו את אוכלוסיית היעד, תתי-קבוצות חשובות, סביבה קלינית ושימוש על ידי צוות האדם-בינה מלאכותית, תשומות מדידה וגורמים מתערבים (confounders) פוטנציאליים[2].

9. **העברת מידע חיוני למשתמשים בצורה שקופה וברורה.** מתן גישה למשתמשים למידע ברור בהקשר רלבנטי, ואשר מותאם לאוכלוסיית היעד (כגון: ספקי שירות או מטופלים), במטרה לבסס את האוטונומיה של מקבל ההחלטות, וכן את אמון המשתמשים במודל. מידע זה יכלול את עצם השימוש במודל למידת מכונה (גילוי), וכן את השימוש המיועד של המוצר והתוויות לשימוש, ביצועי המודל עבור תתי-קבוצות רלבנטיות, מאפייני הנתונים אשר שימשו לאימון ובדיקת המודל, סטנדרטים ופרוטוקולים מקצועיים ששימשו להנחות המודל, הנתונים הנדרשים להזנה, מגבלות ידועות במודל, פרשנות ממשק משתמש, הסבר מותאם על ממצאי המודל(explainability), אינטגרציה של המודל בתהליך העבודה הקלינית, כולל תהליך האימות (validation) הנדרש לפני שימוש ראשוני, יידוע המשתמשים בשינויים ועדכונים במכשיר שמקורם מניטור ביצועים בעולם האמיתי (Real World Evidence), אמצעים לתקשור חששות לגבי המוצר לגורם המפתח.

10. **ניטור ביצועי המודלים המוטמעים וניהול הסיכונים הכרוכים באימון מחדש** בעולם האמיתי, תוך התמקדות בשימור או שיפור במדדי הבטיחות ובביצועים. בנוסף, כאשר מודלים מאומנים מעת לעת או באופן מתמשך לאחר ההטמעה, קיימות בקרות מתאימות לניהול סיכונים של התאמת יתר, הטיה לא מכוונת או ירידה בביצועי המודל (לדוגמה, סחף של מערכי נתונים[3]), אשר עלולה להשפיע על הבטיחות והביצועים של המודל כפי שהוא משמש את צוות האדם-בינה מלאכותית. יש מקום לבחון במסגרת ניהול הסיכונים את תוכנית ההטמעה, גם בהינתן הפעלה בסביבות שונות.

---

[2] Confounders בהקשר זה הינם משתנים מתערבים שהקשר שלהם למשתנה המטרה אינו אמין ואינו ניתן להכללה על בסיסי מידע חדשים. להרחבה על מהות היבט זה נמליץ על הפסקה " Accidentally fitting confounders versus true signal" במאמר שבקישור הבא: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6821018/

[3] Data drift – תופעה של שינויים בהתפלגות המידע