



משרד מבקר המדינה
ונציב תלונות הציבור

מראות

גיליון מס' 66 | יוני 2026



עיונים בביקורת המדינה ובבירור תלונות הציבור



משרד מבקר המדינה
ונציב תלונות הציבור

מראות

עיונים בביקורת המדינה ובבירור תלונות הציבור

גיליון מס' 66

ירושלים, תמוז התשפ"ו, יוני 2026

יו"ר המערכת והעורך הראשי: מר צביקה אוהל

עורכת המשנה: גב' טלי כהן

צוות המערכת: ד"ר האלה אבו חלה, מר תמיר גזיאל, ד"ר אביחי לויט,

מר אילן מוריץ, גב' אהובה נבו, גב' מאיה נוי, גב' מיכל פריאל

עורכת לשון אחראית: גב' גיתית עטר

עורכות הלשון: גב' טל רייך, גב' דקלה מארק אופנהיימר, גב' נעה רוזן,

גב' חוה לוסטיג-מלימובקה

מעצבת ראשית: גב' הדר קאפח־סייג

עיצוב גרפי: גב' שיראל אביטל

מספר קטלוגי: 0793-194 | ISSN: 2026-K-001

התמונות במגזין זה נוצרו בבינה מלאכותית (AI) באמצעות פלטפורמות שונות.

המאמרים המתפרסמים בגיליון זה הם על דעת הכותבים ובאחריותם בלבד. אין האמור בהם משקף את עמדתו של מבקר המדינה ונציב תלונות הציבור בכל סוגיה, ואין הוא מחייב את מבקר המדינה ונציב תלונות הציבור או את משרדו.

תוכן העניינים

6	דבר מבקר המדינה ונציב תלונות הציבור	
8	דבר העורך	
	שימוש בכלי AI בביקורת המדינה: מסגרת מושגית-נורמטיבית להליך ביקורת מבוסס ראיות	
10	מיקו רותם	
	סוכני בינה מלאכותית בשירות ביקורת המדינה	
26	חגי לדרר ואופיר מזרחי	
	השלם גדול מסך חלקיו: מטא-ביקורת בשלטון המקומי בעידן הבינה המלאכותית	
38	אילת רייכמן-עמית	
	הבינה המלאכותית בשירות הביקורת הכלכלית: ניתוח אוטומטי של נתונים כלכליים	
58	האלה אבו-חלה ואופיר מזרחי	
	ביקורת סייבר מונחית מדדים	
74	גיל גזית	
	איומי סייבר על מערכות ביטחוניות בלתי מאוישות בעידן הבינה המלאכותית - הצעה למסגרת ביקורת מעודכנת להערכה ולבקרה	
82	איתן דהן, משה מורנו, אבישי יהושע, אורן בר	
	המבקר האחרון? בינה מלאכותית ועתיד מקצוע הביקורת	קולות מהשטח
96	פבל קריינין	
	ישראל כמובילת סייבר עולמית - הדרך, ההובלה וחיוניות הביקורת להמשך ההצלחה	
116	יגאל אונא	
	ביקורת המדינה על שימוש בבינה מלאכותית בגופים ציבוריים: עקרונות לבחינה מוסדית ונורמטיבית	קולות מהאקדמיה
126	יובל ריינפלד וענבל בן ארצי	
	ביקורת על מערכות בינה מלאכותית - מדריך למבקרי המגזר הציבורי (מתורגם)	קולות מהעולם
140	מוסדות הביקורת העליונים של ברזיל, פינלנד, גרמניה, הולנד, נורווגיה ובריטניה	
	בינה מלאכותית וביקורת ציבורית: אתגרים והזדמנויות עבור מוסדות ביקורת עליונים (מתורגם)	
160	דולורס גנארו-מויה, אנטוניו מנאל לופז-הרננדז ומריה גודז	

דבר מבקר המדינה ונציב תלונות הציבור

על פערים מהותיים. הממצאים והפערים שהועלו מאשרים את הנחת היסוד ולפיה בסביבה האסטרטגית העכשווית פרצות וחולשות במרחב הסייבר הן ליקוי קריטי.

גיליון זה מספק נקודת מבט ייחודית לגבי התפקיד המתפתח של ביקורת המדינה ונציבות תלונות הציבור והאתגרים העומדים לפניו, ופרסומו חופף לסיום כהונתי כמבקר המדינה ונציב תלונות הציבור. המאמרים השונים בוחנים את המעבר מתיאוריה למעשה בשילוב כלי AI בעבודת הביקורת - החל מניתוח דוחות כספיים, דרך גיבוש מתודולוגיות וכלים יישומיים שיסייעו להגביר את יעילותה ודיוקה של הביקורת וכלה ברעיון של מטא-ביקורת בשלטון המקומי, הבוחנת את התמונה הגדולה והמערכתית. חלק מהמאמרים בגיליון זה עוסקים בהתאמות הנדרשות בעולם הביקורת, בין היתר כדי לבחון מערכות ביטחוניות בלתי מאוישות, להתמודד עם אתגרי הביקורת מונחית המדדים בתחום הגנת הסייבר ולוודא כי מעמדה של ישראל כמובילת סייבר עולמית נשמר ומתחזק. הביקורת חיונית כדי להבטיח שהטכנולוגיות המתקדמות ביותר יהיו גם בטוחות, מוגנות וחסיונות. מעל לכול, עלינו לוודא שהשימוש הגובר בבניה מלאכותית במגזר הציבורי נעשה תוך שמירה על עקרונות הממשל התקין. בגיליון זה אנו מציגים עקרונות מנחים לבחינה מוסדית ונורמטיבית של הטמעת הבניה המלאכותית בגופים ציבוריים. אנו מדגישים את החשיבות של הנשיאה באחריות, השקיפות, הבקרה האנושית ומניעת הפגיעה בזכויות הפרט, כדי שהטכנולוגיה תשרת את האדם, ולא להפך.

המנדט הכפול של משרדי - הן כמבקר המדינה הן כנציב תלונות הציבור - מחייב גישה כוללת המאזנת בין קדמה טכנולוגית לבין הגנה על זכויות הפרט. שילוב הבניה המלאכותית במינהל הציבורי טומן בחובו הבטחה לייעול, אולם גם מציב סיכונים לשקיפות, לאחריותיות ולזכויות אזרחיות בסיסיות.

אני מתכבד להציג לפניכם את הגיליון החדש של "מראות", כתב העת של משרד מבקר המדינה, אשר מוקדש השנה לצומת המרתק והמאתגר שבו נפגשים עולמות הביקורת, הבניה המלאכותית (AI) וההגנה בסייבר.

המהפכה הטכנולוגית מאיצה, והבניה המלאכותית הופכת מחזון למציאות המשפיעה על כל היבט בחיינו, ובכלל זה על המגזר הציבורי. כגוף הביקורת ובירור תלונות הציבור העליון של מדינת ישראל, עלינו לא רק להתאים את עצמנו לעידן החדש, אלא להבטיח שהקדמה הטכנולוגית תשרת את כלל אזרחי המדינה באופן מיטבי, יעיל, שקוף ומוסרי. אנו שואפים להבין כיצד הופכים את "הקופסה השחורה" של הבניה המלאכותית לכלי עבודה שקוף ואפקטיבי על שולחן מבקר המדינה ונציב תלונות הציבור. לצד ההזדמנויות, אנו נדרשים לייחד תשומת לב לאתגרים החדשים שהטכנולוגיה מציבה.

עם כניסתי לתפקיד בשנת 2019 הגדרתי את תחום הסייבר ומערכות המידע כאחד מנושאי הליבה שבהם תעסוק ביקורת המדינה. לנגד עיני עמד באופן ברור הצורך לבחון את מוכנות הגופים הציבוריים לאתגרי העולם הדיגיטלי, להבטיח את ההגנה על תשתיות קריטיות ועל המידע האישי של כל אזרח ולוודא שמדינת ישראל לא רק תתמודד עם האיומים, אלא גם תממש את ההזדמנויות הגלומות במהפכת הבניה המלאכותית. דאגתי להוביל קו ביקורת הדורש מהממשלה לאמץ אסטרטגיה לאומית ארוכת טווח בתחום הבניה המלאכותית, והתרעתי כי היעדר מדיניות סדורה עלול לגרום לישראל לאבד את מעמדה כמובילה עולמית בתחום זה.

השנים האחרונות הדגישו את תלותה של ישראל במערכות טכנולוגיות ואישישה את הרלוונטיות של ממצאי דוחות ביקורת רבים שפורסמו בשנים אלו. דוחות מרכזיים שעסקו בהגנת הסייבר על מוסדות לאומיים, ביעילותן של מערכות התרעה אזרחיות ובאבטחת הגופים הביטחוניים הציבו

כתב עת זה הוא כלי הצופה פני עתיד בנוגע לעיצוב פניה של ביקורת המגזר הציבורי. האתגרים הטכנולוגיים והאתיים שמציבות הבינה המלאכותית והלוחמה הקיברנטית יסיפו להתעצם. משרד מבקר המדינה ונציב תלונות הציבור מתחייב להוסיף לעמוד בחזית התפתחויות אלה, תוך חידוד מתמיד של מתודולוגיות הביקורת שלו והבטחה שהממשל בישראל יתפתח באופן בטוח, חדשני וערכי.

לסיים, אבקש להביע את תודתי העמוקה לכל הכותבות והכותבים אשר תרמו מפרי עטם ומידיעותיהם להעשרתו של גיליון חשוב זה. תודה מיוחדת שלוחה לחברי המערכת בראשות מנכ"ל המשרד מר ישי וקנין, אשר עמלו במסירות על איסוף המאמרים, עריכתם והבאת כתב העת לדפוס.

אני סמוך ובטוח כי המאמרים בגיליון זה מרימים תרומה משמעותית לשיח המקצועי ומספקים כלים והשראה לעוסקים במלאכת הביקורת ובמינהל הציבורי. גיליון זה מוגש לכם מתוך תחושת אחריות מוסדית עמוקה ואמונה איתנה בחשיבותה המתמדת של ביקורת המדינה לעתיד האומה.

אני מזמין אתכם לקריאה מעמיקה ומעוררת מחשבה.



מתניהו אנגלמן

מבקר המדינה
ונציב תלונות הציבור

דבר העורך

העבודה, ובהם סוכן בינה מלאכותית המבוסס על המדריך המקצועי לביקורת המדינה וכלי AI המשמשים לניתוח דוחות כספיים. המאמר השלישי: "השלם גדול מסך חלקיו: מטא-ביקורת בשלטון המקומי בעידן הבינה המלאכותית" מציג ניתוח מתכלל של ממצאי מאות דוחות ביקורת בשלטון המקומי, במבט רטרוספקטיבי ורפלקטיבי, ככלי אסטרטגי ליעול התכנון של עבודת הביקורת ולטיובה.

המאמרים החותמים את החלק הראשון של הגיליון עוסקים באתגר המורכב של הגנת התשתיות הקריטיות בגופים המבוקרים מפני איומי הסייבר בעידן החדש. המאמר "ביקורת סייבר מונחית מדדים" מבהיר את חשיבותה של ביקורת המדינה בזיהוי פערים ביכולתם של הגופים המאסדרים לנטר סיכונים פוטנציאליים הגלומים בתרחישי הייחוס. כמו כן, במאמר מוצגים כלים המסייעים לגופים אלה בקביעת סדרי עדיפויות לטיפול בפערים במסגרת תהליכי קבלת החלטות מושכלים. המאמר "איומי סייבר על מערכות ביטחוניות בלתי מאוישות בעידן הבינה המלאכותית - הצעה למסגרת ביקורת מעודכנת להערכה ולבקרה" בוחן כיצד יש לעדכן את עבודת הביקורת בתחום הגנת הסייבר לנוכח ההתבססות של מנגנון קבלת ההחלטות - רכיב הליבה של המערכות הביטחוניות הבלתי מאוישות - על בינה מלאכותית ועל מודלים ונתונים דינמיים.

בחלק השני של הגיליון מובאים "קולות מהשטח": אנשים בעלי ניסיון רב בתחומי הבינה המלאכותית והסייבר העלו על הכתב את השקפת עולמם ואת התובנות שהפיקו בתחומים אלה במשך השנים. במישור הלאומי, ממאמרו של ראש מערך הסייבר הלאומי לשעבר, "ישראל כמובילת סייבר עולמית - הדרך לשם, ביטוי ההובלה בפועל וחיוניות הביקורת להמשך ההצלחה", עולה כי שילוב הדוק בין הגנת סייבר מקצועית לביקורת עצמאית ושיטתית הוא תנאי חיוני לשימור יתרונה היחסי של ישראל בתחום הסייבר ולהבטחת עמידותה בפני האיומים הנשקפים במרחב הדיגיטלי.

המאמר הבא בחלק זה, "המבקר האחרון? בינה מלאכותית ועתיד מקצוע הביקורת" הוא אחד ממאמרי המפתח של

המפגש בין בינה מלאכותית (AI), הגנת הסייבר, המגזר הציבורי ותחום ביקורת המדינה ובירור תלונות הציבור אינו עוד פרק באבולוציה הטכנולוגית, אלא נקודת מפנה מתודולוגית ורעיונית. אם בעבר הסתמכו צוותי הביקורת ומבררי התלונות בעיקר על תהליכי קבלת החלטות המבוססים על נורמות כתובות ורשומות סטטיות, הרי שכיום הם ניצבים מול מערכות לומדות ומשתנות הפועלות במהירות מסחררת. טכנולוגיות אלו משנות מן היסוד הן את מושאי הביקורת - הגופים המבוקרים הנעזרים בבינה מלאכותית והחשופים לאיומי סייבר - והן את כלי הביקורת עצמם.

בגיליון מיוחד זה של כתב העת "מראות" אנו מבקשים להניח לרגע בצד את השיח הנלהב המלווה את עולם ה-AI שכן ברצוננו להתמקד בהיבטים המעשיים של השימוש בבינה מלאכותית ולהציע למנהלי הביקורת, למבררי התלונות, למבקרי הפנים ולכלל העוסקים בתחום הביקורת מעטפת מקצועית, המאגדת מתודולוגיה סדורה וארגז כלים יישומי.

מבנה כתב העת

בחלקו הראשון של הגיליון עובדי ביקורת ממשרד מבקר המדינה ונציגי תלונות הציבור סוקרים היבטים שונים ביישום ארגז הכלים החדש העומד לרשותם בהווה, תוך מבט צופה פני עתיד על היתרונות והסיכונים הגלומים בו.

המאמר הפותח את הגיליון, "שימוש בכלי AI בביקורת המדינה: מסגרת מושגית-נורמטיבית להליך ביקורת מבוסס ראיות", מציג מתווה המתרגם עקרונות כלליים של שימוש ב-AI לקריטריונים של ביקורת. מתווה זה משרטט מסגרת מושגית-נורמטיבית לשילוב כלי AI בעבודת ביקורת המדינה בלי לפגוע בשיקול הדעת המקצועי-אנושי.

במאמר השני: "פוטנציאל השימוש בסוכני בינה מלאכותית (AI AGENT) המבוססים על הנחיות ומדריכים פנים ארגוניים בעבודת ביקורת המדינה" ובמאמר הרביעי: "הבינה המלאכותית בשירות הביקורת הכלכלית: ניתוח אוטומטי של נתונים כלכליים" נבחנים פיתוחים המשנים את אופי

לא רק תתמודד עם הסיכונים הכרוכים במהפכת הבינה המלאכותית, אלא גם תנצל את ההזדמנויות הגלומות בה.

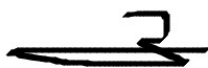
נוסף על כך, גיליון זה של "מראות" עולה בקנה אחד תפיסתו המקצועית של המבקר, הגורסת כי יש לפעול להפיכת "הקופסה השחורה" של הבינה המלאכותית לכלי עבודה שקוף, אפקטיבי ונגיש עבור צוותי הביקורת ומבררי התלונות, לצד מתן תשומת לב לאתגרים החדשים הכרוכים במהפכה טכנולוגית זו.

אני מבקש להביע הוקרה עמוקה כלפי מנכ"ל המשרד, מר ישי וקנין, אשר הוביל את מהפכת הטמעתה של הבינה המלאכותית בין כותלי הארגון והפך אותה מחזון רחוק לפרקטיקה חיה בעבודתנו. תודה אישית על האמון המלא שנתת בי עת מינית אותי לתפקיד עורך הגיליון.

תודה מיוחדת לכל כותבי המאמרים בגיליון זה - עובדות ועובדים במשרד מבקר המדינה ונציגי תלונות הציבור, אנשי אקדמיה ואנשי מקצוע מהשורה הראשונה - כולם עוסקים בתחומי הבינה המלאכותית והסייבר - על שנענו לבקשתנו ונרתמו ברצון לכתיבת המאמרים ולשיתוף הידע והניסיון המקצועי שצברו.

לעבודה המאומצת של הכנת הגיליון היו שותפים רבים: תודה מקרב לב לעורכת המשנה, גב' טלי כהן, על הניהול המקצועי, העין החדה והתמיכה לאורך הדרך. תודה לצוות המערכת המסור על העבודה הבלתי פוסקת מאחורי הקלעים: גב' אהובה נבו, גב' מאיה נוי, גב' מיכל פריאל, ד"ר האלה אבו חלה, ד"ר אביחי לויט, מר תמיר גזיאל ומר אילן מוריץ. תודה לצוות עריכת הלשון בראשות גב' גיתית עטר ולצוות העיצוב הגרפי - גב' הדן קאפח סייג וגב' שיראל אורנלה אביטל, שהפכו אסופת מאמרים מורכבת לתוצר בהיר, קריא ומהודר, העומד בסטנדרטים המקצועיים הגבוהים ביותר.

אנו מאמינים כי אסופת המאמרים שלפניכם תתרום תרומה של ממש לעיצוב פניה של ביקורת המדינה ולהתאמת כלי הביקורת לאתגרי העתיד. נאחל לכם קריאה מעשירה ומעוררת מחשבה.



צביקה אהל

עורך הגיליון ומערכת כתב העת

הגיליון. ייחודו של המאמר אינו במיפוי הטכנולוגיה, כי אם בהיפוך שאלת המוצא. בעוד הדיון הציבורי מתמקד בשאלה כיצד אפשר לשלב את הבינה המלאכותית בפעילותו של המבקר האנושי כדי לשפר את תוצרי עבודתו, המאמר מבהיר כי זוהי שאלה זהירה מדי, כמעט מתחמקת, וכי יש להעז ולהציג שאלה מאתגרת יותר, כיצד המבקר האנושי ישתלב בעידן שכולו מנוהל על ידי בינה מלאכותית. מוסדות הביקורת שיקדימו להיערך למהפכה זו לא זו בלבד שישמרו על הרלוונטיות שלהם אלא שהם יתוו מחדש את אמות המידה שלפיהן יתמדד ביקורת המדינה בעשורים הבאים.

בחלק השלישי מאוגדים מאמרים של חוקרים מהאקדמיה בארץ ובעולם. במאמר הראשון, "ביקורת המדינה על שימוש בבינה מלאכותית בגופים ציבוריים: עקרונות לבחינה מוסדית ונורמטיבית" מובהר כי בהיעדר מסגרת חקיקתית רוחבית, ביקורת המדינה עשויה לחרוג מהתפקיד המסורתי של חשיפת ליקויים ולסייע בעיצוב נורמות ציבוריות לשימוש אחראי בבינה מלאכותית. לשם כך נדרשת הקמת תשתית ביקורתית חדשה, המתבססת על שפה מקצועית עדכנית, מומחיות טכנולוגית, שיתוף פעולה בין-מוסדי, גיבוש הנחיות מקדימות למבוקרים ויכולת למידה מן המתרחש בתחום זה בזירה הבין-לאומית.

המאמר השני, של חוקרים מספרד, "בינה מלאכותית וביקורת ציבורית: אתגרים והזדמנויות עבור מוסדות ביקורת עליונים", מנתח את תוצאות הטמעתה של הבינה המלאכותית במגזר הציבורי וכן את האתגרים שהתפתחויות טכנולוגיות אלה מציבות לפני מוסדות ביקורת במסגרת מאמציהם לשיפור האפקטיביות והיעילות של עבודתם.

בסיפה של הגיליון מובא מדריך למבקר המגזר הציבורי העושים ביקורת על מערכות בינה מלאכותית. המדריך, שגובש במוסדות הביקורת העליונים של ברזיל, פינלנד, גרמניה, הולנד, נורווגיה ובריטניה, תורגם לעברית כדי שכלל העוסקים בנושא בארץ יוכלו להסתייע בו.

מילות תודה והערכה

תודה מיוחדת ועמוקה נתונה למבקר המדינה ונציגי תלונות הציבור, מר מתניהו אנגלמן, המסיים בימים אלו את תפקידו. במהלך כהונתו השכיל המבקר לזהות שהבינה המלאכותית הופכת מחזון למציאות ומשפיעה על כל היבט בחיינו, ובכלל זה על המגזר הציבורי. תובנה זו חידדה את הצורך לבחון את מוכנות הגופים הציבוריים לאתגרי העידן הדיגיטלי, להבטיח את ההגנה על תשתיות קריטיות ועל המידע האישי של כל אזרח ולהבטיח שמדינת ישראל

10

שימוש בכלי AI בביקורת המדינה: מסגרת מושגית- נורמטיבית להליך ביקורת מבוסס ראיות

מיקו רותם

עו"ד מיקו רותם הוא מנהל ביקורת בכיר באגף לביקורת תחומי האכיפה והמשפט בחטיבה לביקורת משרדי הממשלה ומוסדות השלטון שבמשרד מבקר המדינה ונציב תלונות הציבור.

המאמר בקצרה

- השיח הקיים בנושא שימוש בכלי AI בגופי ממשל מציע עקרונות חשובים של שקיפות, אחריותיות ופיקוח אנושי, אך אינו מספיק כדי להסביר כיצד אפשר לשלב תוצרי AI בשרשרת הביקורת של נורמה - ראייה - ממצא - מסקנה - המלצה.
- שימוש ב-AI בביקורת המדינה אינו רק שאלה של יעילות, אלא שאלת תוקף ראייתי והצדקה של ממצאים.
- לצורך שימוש תקף ב-AI בביקורת המדינה נדרשת הגדרת "שכבת תרגום" מושגית-נורמטיבית בין עקרונות של ממשל AI לבין דרישות ההוכחה וההצדקה של ביקורת המדינה.
- המאמר מציע שכבת תרגום שממירה עקרונות כלליים לשימוש ב-AI לקריטריוני ביקורת וכן מפתח מסגרת מושגית-נורמטיבית לשילוב כלי AI בעבודת ביקורת המדינה בלי לעקוף את שיקול הדעת המקצועי-אנושי.

מבוא | הגדרת הפער | ייחודיות מתודולוגית של ביקורת המדינה | ממשל AI | אחריותיות אלגוריתמית בהקשר של עבודת הביקורת | נורמות, ראיות ושימוש ב-AI | תחום התחולה של השימוש ב-AI בביקורת המדינה | הבחנות מושגיות | סיכונים, מגבלות אפיסטמיות ושרשרת הראיות הביקורתית | מקומם של הפיקוח האנושי ושיקול הדעת המקצועי | סינתזה: מסגרת מושגית-נורמטיבית לשימוש תקף ב-AI בתוך המערכת הפורמלית של ביקורת המדינה | סיכום

ביקורת המדינה כאשר היא עצמה נעזרת בכלי AI.

המאמר אינו עוסק ב-AI כשלעצמו, ואינו עוסק בביקורת על מערכות AI של גופים מבוקרים. עניינו בשימוש של ביקורת המדינה עצמה בכלי AI בתהליך העבודה הפורמלי של הביקורת, ובתנאים שבהם שימוש כזה יכול להיות תקף מבחינת דרישות ביקורת המדינה. אין מדובר כאן בדיון כללי על חדשנות טכנולוגית במגזר הציבורי, אלא בבחינת מקומו האפשרי של AI בפרקטיקה מוסדית שמטרתה להפיק קביעות ביקורתיות החשופות לבחינה מקצועית, ציבורית ופרלמנטרית.

על יסוד זה, שאלת המחקר שעליה מבקש המאמר לענות היא: כיצד ניתן לשלב שימוש בכלי AI בהליך הפורמלי של ביקורת המדינה בלי לפגוע בתוקף הראייתי ובלוגיקת ההוכחה של הממצאים, המסקנות וההמלצות? הטענה המרכזית של המאמר היא כי שיח ה-AI הקיים, חשוב ככל שיהיה, אינו מספיק לצורך זה. מן הפער הזה נובע הצורך במבנה מעבר "יעודי" - שכבת תרגום מושגית-נורמטיבית - שתתווך בין עקרונות ממשל AI לבין דרישות הנורמה, הראיה, הממצא, המסקנה וההמלצה של ביקורת המדינה. שכבת תרגום זו נדרשת לא כעיקרון נוסף לשיח ה-AI, אלא היא הכרח הנובע מהפער האמור שבין עקרונות כלליים לבין דרישות לוגיקת ההסקה של ביקורת המדינה.

בכך מבקש המאמר לתרום בשלושה מישורים משולבים: במישור המושגי - להבהיר את ההבחנה בין שיח ה-AI לבין לוגיקת הביקורת ולפתח שכבת תרגום ביניהם; במישור האפיסטמי - להגדיר את תנאי המעבר התקפים מתוצר אנליטי אל מערך ראיות ביקורת שעל בסיסו בלבד

כניסתם המהירה של כלי AI למרחב העבודה המוסדי ולמגזר הציבורי מעוררת שאלות חדשות גם על אופן השימוש בהם בשדה ביקורת המדינה. בהקשר של ביקורת המדינה השאלה המרכזית איננה אם AI הוא כלי יעיל או ראוי כשלעצמו, אלא אם וכיצד אפשר לשלב בהליך ביקורת מבוסס ראיות בלי לפגוע בתוקף הראייתי ובלוגיקת ההוכחה של הממצאים, המסקנות וההמלצות. נקודת המוצא של מאמר זה היא כי ביקורת המדינה פועלת כמערכת פורמלית של שרשרת היסק לוגי הבנויה מנורמה - ראיה - ממצא - מסקנה - המלצה, ולכן כל דיון בשילוב AI חייב להיבחן מתוך מבנה זה.

הקושי נובע מכך שהשיח המתפתח על AI governance (להלן - ממשל AI), וכן על AI/auditworthy AI/responsible AI (להלן - AI אמין/אחראי) ואחריות אלגוריתמית², מספק עקרונות, מנגנוני בקרה ודרישות של שקיפות, פיקוח, תיעוד ואחריות; אך אינו מספק כשהוא לעצמו שכבת תרגום³ מספקת ללוגיקת העבודה וההוכחה של ביקורת המדינה (Audit logic)⁴. שיח זה עוסק בעיקר בעקרונות, בתכונות רצויות של מערכות ובתנאים לשימוש אחראי ב-AI; ואילו ביקורת המדינה פועלת בלוגיקה פורמלית של נורמה - ראיה - ממצא - מסקנה - המלצה. משום כך השאלה המכריעה איננה רק מה ראוי שיתקיים בשימוש מוסדי ב-AI, אלא כיצד אפשר לבסס, לאמת ולהצדיק טענה ביקורתית במסגרת דוח המבוסס על נורמות וראיות. לכן הפער שבו עוסק המאמר אינו רק פער בין עקרונות לבין יישומם, אלא פער בין שיח על עקרונות ומאפיינים של מערכות לבין לוגיקת ההוכחה של

1 ראו למשל משרד מבקר המדינה ונציב תלונות הציבור המדריך לביקורת המדינה, <https://www.mevaker.gov.il/state-audit/roles-work/state-audit-guide>, כתיבת דוח הביקורת, הנחיות כלליות לכתיבת הדוח ומאפייני הדוח, בעמ' 2: "על צוות הביקורת להקפיד על אלה: כתיבה ברצף הגיוני, שבו הממצאים המסקנות וההמלצות קשורים באופן ברור לנורמות [...] המסקנות וההמלצות נובעות באופן הגיוני ממצאי הביקורת" (ההדגשות במקור, מ"ר); וכן ראו שם, רכיבי דוח הביקורת, בעמ' 8 - 12.

2 AI governance (ממשל AI) הוא מסגרת רחבה שבאמצעותה מוסדות פועלים להסדרת פיתוח, הטמעה, שימוש, בקרה, תיעוד וניהול סיכונים של מערכות AI. הוא כולל קביעת מדיניות, ניהול סיכונים, פיקוח אנושי, שקיפות, אחריות ומנגנוני תיקון, כדי להבטיח שהשימוש ב-AI יתיישב עם הדיון, עם ערכי הארגון ועם ייעודו הציבורי. במאמר זה המונח מציין מסגרת נורמטיבית כללית, המובחנת מלוגיקת העבודה וההוכחה של ביקורת המדינה.

3 Trustworthy AI/Responsible AI (אמין/אחראי) הם מונחי מסגרת המתייחסים ל-AI אמין ו-AI אחראי, כלומר למערכות AI שאמורות לעמוד בדרישות כגון בטיחות, אמינות, שקיפות, פיקוח אנושי ואחריות. במאמר זה הם משמשים רקע נורמטיבי רחב ולא מסגרת מספקת לביקורת המדינה.

4 אחריותיות אלגוריתמית (Algorithmic Accountability) היא החובה לתת דין וחשבון, הסבר והצדקה לאופן השימוש במערכת AI, לתהליך שנוצר באמצעותה ולתוצרים שנגזרו ממנה. במאמר זה היא מובחנת מבוסס ראיות ואינה שקולה לו. יש להדגיש גם כי אחריותיות איננה אחריות (Responsibility), והאחרונה עניינה חובות הביצוע והניהול של תהליך העבודה עם AI בהתאם לכללים, להנחיות ולפרקטיקות מוגדרים.

5 שכבת תרגום (Translation Layer) היא מסגרת מושגית-נורמטיבית המתווכת בין שיח ה-AI העקרוני לבין לוגיקת ההוכחה של ביקורת המדינה. תפקידה להמיר עקרונות כלליים לקריטריונים, תנאי בדיקה ודרישות עיון ראיות המתאימים להליך ביקורת מבוסס ראיות. ביאור והרחבה מופיעים בפרקי המאמר ובעיקר בפרק 3.

6 לוגיקת העבודה וההוכחה של הביקורת (Audit logic) היא המבנה הפורמלי של שרשרת ההיסק בביקורת המדינה. במסגרת ביקורת המדינה עוברת מנורמה לראיה, מראיה לממצא, ממנו למסקנה ולבסוף להמלצה. במאמר זה זהו מסגרת התוקף של כל שימוש ב-AI.

בנקודה זו אין מדובר עוד רק בשאלה כללית של מדיניות טכנולוגית, ניהול סיכונים או אתיקה יישומית במינהל הציבורי, אלא בשאלה אם וכיצד אפשר לשלב כלי AI בפרקטיקה מוסדית שמטרתה להפיק קביעות ביקורתיות מבוססות ראיות, החשופות לבחינה מקצועית, ציבורית ופרלמנטרית.⁸

מן הטעם הזה הבעיה המרכזית איננה מחסור בעקרונות או בתיאוריות, אלא מחסור בשכבת תרגום שלהם לעולם ביקורת המדינה. מסגרות ממשל AI יודעות להגדיר, במידה רבה, את מרחב החובות, הסיכונים והדרישות הנורמטיביות: שקיפות, פיקוח אנושי, אחריות, תיעוד, ניהול סיכונים, אמינות והוגנות.⁹ ואולם בביקורת המדינה אין די בעמידה בדרישות הכלליות של שימוש אחראי ב-AI. השימוש בכלי נבחן גם לפי יכולתו להשתלב בשרשרת הביקורת: נורמה, ראייה, ממצא, מסקנה והמלצה. קרי, מעבר לדרישות הכלליות המוטלות על גופים ציבוריים אחרים כחלק מאחריותם הציבורית, הלוגיקה של ביקורת המדינה - שלפיה הנורמה משמשת אמת מידה להשוואה; הראיות הן הבסיס האמפירי להערכת המצב בפועל; הממצא הוא תוצאת ההשוואה בין השניים; המסקנה מפרשת את הממצאים, מסבירה את משמעותם ומנתחת את סיבותיהם והשלכותיהם; וההמלצה צופה פני עתיד ונגזרת מן הממצאים והמסקנות¹⁰ - מחייבת בחינה מחמירה יותר. לפיכך השאלה האמיתית איננה אם AI עשוי לייעל את עבודת הביקורת, אלא אם אפשר למקם אותו בתוך שרשרת הביקורת בלי לפגוע בלוגיקה שעליה נשען דוח הביקורת. זוהי שכבת התרגום החסרה.¹¹ כאן מתגלה הפער התיאורטי שבו עוסק מאמר

אפשר לגבש ממצא, מסקנה והמלצה; ובמישור המוסדי-ביקורתי - למקם את מנהל הביקורת כבעל האחריות המקצועית להערכת הראיות, לאתגורן ולהצדקת מקומן בדוח הביקורת. במובן זה המאמר אינו עוסק ב-AI כשלעצמו, אלא בשאלה עמוקה יותר: כיצד משלבים טכנולוגיה מייצרת תובנות במערכת שמייצרת הוכחות. לעומת מערכות ציבוריות רגילות, שבהן AI נבחן פעמים רבות לפי יעילותו, בביקורת המדינה השאלה המכריעה היא אם אפשר להצדיק באמצעותו טענה בדוח ביקורת מבוסס ראיות, גם כאשר מדובר בתוצרים איכותיים ומדויקים. המסגרת המוצעת נועדה לאפשר שימוש מושכל ב-AI תוך שמירה על תקפות ביקורתית, והיא גם מניחה בסיס עקרונות ותשתית נורמטיבית לפיתוח עתידי של מודל הפעלה מתודולוגי ומסגרת בקרה, ולביצוע התאמות עתידיות שלהם להתפתחויות הטכנולוגיות.

הגדרת הפער בין ממשל AI לבין לוגיקת העבודה וההוכחה של ביקורת המדינה כאשר היא עצמה נעזרת בכלי AI

כניסתם המהירה של כלי AI לשימוש מוסדי יצרה בשנים האחרונות שדה רחב של עקרונות, מסגרות נורמטיביות, כלי הערכה ומנגנוני בקרה לשימוש "אמין", "אחראי" ו"מנוהל היטב" בטכנולוגיות AI.⁷ ואולם התפתחותו של שדה זה עדיין אינה משיבה על השאלה המתעוררת כאשר ביקורת המדינה עצמה מבקשת להיעזר בכלי AI במהלך עבודתה.

7 OECD, *Recommendation of the Council on Artificial Intelligence* (OECD/LEGAL/0449) (2019 amended 2024) <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449>; Regulation (EU) 2024/1689 Laying Down Harmonised Rules on Artificial Intelligence (AI Act), OJ. L of 12 July 2024 (entered into force 1 August 2024), <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>; National Institute of Standards and Technology (NIST), *AI Risk Management Framework (AI RMF 1.0)*, NIST AI 100-1 (2023) <https://www.nist.gov/itl/ai-risk-management-framework>; Luciano Floridi et al., *AI4People: An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*, 28 *Minds & Machines* 689 (2018); Anna Jobin, Marcello Ienca & Effy Vayena, *The Global Landscape of AI Ethics Guidelines*, 1 *Nature Machine Intelligence* 389 (2019).

8 משרד מבקר המדינה **המדריך לביקורת המדינה**, לעיל ה"ש 1, תהליך הביקורת השלם, בעמ' 1 - 4; רכיבי דוח הביקורת: פעולות הביקורת, תמונת מצב העולה מן הביקורת ועיקרי המלצות הביקורת, בעמ' 3 - 5.

9 OECD, *Recommendation*, לעיל ה"ש 7; AI Risk; National Institute of Standards and Technology (NIST), לעיל ה"ש 7.

10 משרד מבקר המדינה **המדריך לביקורת המדינה**, לעיל ה"ש 1, ראיות הביקורת: מהי ראיית ביקורת? וכן הערכת ראיות ביקורת, בעמ' 1 - 2; שם, גיבוש ממצא ביקורת: מהו ממצא ביקורת?, בעמ' 1 - 2; שם, גיבוש מסקנת ביקורת: מהי מסקנת הביקורת? וכן רכיבי מסקנת הביקורת, בעמ' 1 - 2; שם, גיבוש המלצת ביקורת: מהי המלצת ביקורת?, בעמ' 1 - 2.

11 Jakob Møkander & Luciano Floridi, *Ethics-Based Auditing to Develop Trustworthy AI*, 31 *Minds & Machines* 323 (2021); International Organization of Supreme Audit Institution (INTOSAI), *ISSAI 100: Fundamental Principles of Public-Sector Auditing* (2019) <https://www.issai.org/wp-content/uploads/2019/08/ISSAI-100-Fundamental-Principles-of-Public-Sector-Auditing-1.pdf>; International Organization of Supreme Audit Institution (INTOSAI), *ISSAI 300: Performance Audit Principles* (2019), <https://www.issai.org/wp-content/uploads/2019/08/ISSAI-300-Performance-Audit-Principles.pdf>; International Organization of Supreme Audit Institution (INTOSAI), *ISSAI 400: Compliance Audit Principles* (2019), <https://www.issai.org/wp-content/uploads/2019/08/ISSAI-400.pdf>

זה. שיח ממשל AI קובע מה אמור להתקיים; לוגיקת הביקורת שואלת כיצד אפשר לתקף ולאמת ממצא ביקורת, באילו ראיות, על בסיס אילו קריטריונים, ובאיזו דרגת ביטחון אפשר לנסח ממצא ומסקנה¹². במובן זה, הפער בין ממשל AI לבין ביקורת המדינה הוא פער בין מסגרת כללית המסדירה שימוש אחראי בכלי AI במגזר הציבורי לבין לוגיקת ההוכחה של ביקורת המדינה כאשר היא עצמה נעזרת בכלים אלה. לכן אין די בכך שהשימוש בכלי יהיה אחראי, שקוף או מפקח; יש להראות כיצד אפשר לשלב בשרשרת הביקורת בלי לפגוע בדרישות הנורמה, הראיה, הממצא, המסקנה וההמלצה¹³. מכאן שכל ניסיון לשלב AI בביקורת המדינה מחייב מעבר מן השאלה "מה ראוי" אל השאלה "מה אפשר לבסס, לאמת ולהצדיק במסגרת דוח ביקורת המבוסס על ראיות".

השלכה ישירה של הבחנה זו היא ההבחנה בין תוצר AI¹⁴ לבין ראיית ביקורת. תוצר של כלי AI עשוי להיות מועיל מבחינה אנליטית: הוא יכול לסייע באיתור כיווני בדיקה, בסיווג חומרים, בארגון נתונים, בזיהוי דפוסים או בניסוח ראשוני של תובנות. אך תועלת אנליטית אינה שקולה למעמד ראייתי. ראיית ביקורת היא מידע המשמש את צוות הביקורת כדי לתקף ולבסס ממצאים, ולאפשר הסקת מסקנות ברמת ודאות מספקת; עליה להיות רלוונטית, תקפה, מהימנה ומספיקה. ממצא ביקורת הוא מידע עובדתי רלוונטי הנובע מהשוואה בין הנורמה לבין המצב בפועל כפי שהוא עולה מראיות הביקורת¹⁵. מכאן שתוצר

AI אינו ראייה מעצם הפקתו, וממילא אינו יכול לשמש ממצא מעצם הופעתו. לכל היותר, לאחר עיבוד ביקורתי, אימות, הצלבה ומיקום ביחס לנורמה, אפשר לבחון אם וכיצד תוצר ה-AI עשוי להשתלב במערך ראיות הביקורת. לכן המעבר התקף אינו תוצר AI ← ממצא, אלא: תוצר AI ← עיבוד ביקורתי ← ראיות ← ממצא¹⁶.

הבחנה זו מתחדדת עוד יותר לנוכח המגבלות האפיסטמיות של כלי AI, ובפרט של מערכות גנרטיביות. הספרות מלמדת כי אי-ודאות היא תכונה מבנית של מערכות למידה, וכי Hallucination (להלן - הזיה) אינה תקלה שולית בלבד, אלא דפוס כשל מובנה המחייב מנגנוני אימות, ריסון והימנעות מהסקה ישירה¹⁷. בהקשר של ביקורת המדינה משמעות הדבר איננה רק צורך טכני בבקרה נוספת, אלא צורך מושגי בהבחנה חדה בין פלט אנליטי או הסתברותי לבין תשתית ראייתית שאפשר לבסס עליה ממצא בדוח ביקורת. דווקא משום שביקורת המדינה מציגה ליקויים והמלצות במגזר הציבורי, לא די שפלט של AI יהיה "משכנע" או "מנסח נכון", ואף לא רק "מדויק" לפי עקרונות ממשל AI; גם לאחר בדיקה, הצלבה ושיקול דעת ביקורתי השאלה היא רק אם וכיצד אפשר לשלב בתשתית ראיות הביקורת, ולא אם הוקנה לו מעמד ראייתי עצמאי¹⁸.

מכאן נובע גם אופיו של המהלך המוצע במאמר: הבחנה שיטתית בין שיח ממשל AI לבין לוגיקת הביקורת, העברת מוקד הדיון מן הראוי או היעיל אל הניתן לביסוס במסגרת דוח ביקורת המדינה, ומיקום השימוש ב-AI בשרשרת

Brent Mittelstadt, *Principles Alone Cannot Guarantee Ethical AI*, 1 Nature Machine Intelligence 501 (2019); Jessica Morley et al., *From 12 What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices*, 26 Science and Engineering Ethics 2141 (2020).

Artificial Intelligence and Public, Dolores Genaro-Moya, Antonio Manuel López-Hernández & Mariia Godz, 11; INTOSAI, *ISSAI 100 13 Sector Auditing: Challenges and Opportunities for Supreme Audit Institutions*, 6 WORLD 78 (2025).

14 **תוצר AI** הוא פלט המופק על ידי מערכת בינה מלאכותית, ובכלל זה תוצרי ניתוח טקסטואלי, ניתוח נתונים או ניתוח מערכות מידע, המשקף עיבוד הסתברותי של מידע, אך הוא כשלעצמו אינו ראיה, ממצא או מסקנה במסגרת הליך ביקורת המדינה

15 משרד מבקר המדינה **המדריך לביקורת המדינה**, לעיל ה"ש 1, ראיות הביקורת: הערכת ראיות ביקורת, בעמ' 1 - 2; שם, גיבוש ממצא ביקורת: מהו ממצא ביקורת?, בעמ' 1 - 2.

16 שם, לעיל ה"ש 15; שם, לעיל ה"ש 1, תהליך הביקורת השלם, בעמ' 3 - 4.

17 **Hallucination (הזיה)** היא תופעה שבה מערכת AI מייצרת פלט שגוי, לא מבוסס או מטעה, לעיתים בניסוח בטוח ומשכנע. במאמר זה המונח משמש לציון מגבלה אפיסטמית המחייבת אימות, הצלבה והימנעות מהסקה ישירה.

18 Eyke Hüllermeier & Willem Waegeman, *Aleatoric and Epistemic Uncertainty in Machine Learning: An Introduction to Concepts and Methods*, 110 Machine Learning 457 (2021); Lei Huang et al., *A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions*, 43 ACM Transactions On Information Systems 1 (2025).

19 משרד מבקר המדינה **המדריך לביקורת המדינה**, לעיל ה"ש 1, ראיות הביקורת: הערכת ראיות הביקורת, בעמ' 1 - 2; INTOSAI, *ISSAI 100*, לעיל ה"ש 11.

אמון הציבור בעבודת משרד מבקר המדינה, באובייקטיביות של הדוחות ובאיכותם המקצועית²⁴.

גם המבנה הפנימי של הדוח מבטא ייחוד מתודולוגי זה. ממצא ביקורת הוא "לב ליבו" של דוח הביקורת, אך הוא איננו נקודת הסיום. המסקנה מסתמכת על הממצאים, מבהירה אותם, מנתחת את סיבותיהם, תוצאותיהם והשלכותיהם, ואינה חזרה גרידא עליהם. ההמלצה, בתורה, היא פעולה מתקנת צופה פני עתיד, שיש בה ערך מוסף, ואשר אמורה לנבוע מן הממצאים והמסקנות ולתת, ככל האפשר, מענה לסיבות הפער²⁵. השרשרת הזאת איננה רק סדר עריכתו; היא מבנה הצדקה ביקורתי הבוחן אם ממצא, מסקנה או המלצה בוססו כדון על נורמה וראיות. כל חוליה תלויה בזו שקדמה לה, וכל כשל באחת מהן מקרין על תוקף החוליות הבאות. לכן תובנה אנליטית אינה ממצא; ממצא אינו מסקנה; ומסקנה אינה המלצה. ממילא גם פלט של AI אינו יכול לדלג על חוליות אלו ולשמש להן תחליף.

מכאן נובעת גם חובת הזהירות המיוחדת של ביקורת המדינה. מאחר שהדוח מיועד הן לגוף המבוקר, הן לציבור והן לפיקוח הפרלמנטרי, אין די בכך שטענה ביקורתית תהיה סבירה או אינטואיטיבית. עליה להיות ניתנת להצדקה, לבחינה ולשחזור במונחים של נורמות, ראיות, שיקול דעת ותהליך עבודה. אפילו כאשר הביקורת מבקשת להצביע על תוצאה אפשרית או על סיכון פוטנציאלי, ההנחיות מחייבות זהירות יתרה וביסוס של הקשר בין הפער לבין התוצאה האפשרית²⁶.

חובה זו מחדדת את ההבדל בין ביקורת המדינה לבין מסגרת כללית של ממשל AI: ביקורת המדינה איננה מסתפקת בהכונה ובבקרה; היא מחייבת שרשרת הצדקה כתובה ומובנית. כאן מתחדד גם ההבדל בין Accountability (אחריותיות) במובנה הכללי לבין הצדקה ביקורתית: אחריותיות דורשת יכולת לתת דין וחשבון בנוגע לאופן השימוש במערכת AI, לתהליך שנוצר באמצעותה ולתוצרים

הפורמלית של נורמה, ראייה, ממצא, מסקנה והמלצה²⁰. לפיכך השאלה הבאה איננה עוד מה AI יודע ויכול לעשות, אלא מה מייחד את השימוש בכלי AI בביקורת המדינה.

ייחודיות ביקורת המדינה כנקודת מוצא מתודולוגית

ייחודיותה של ביקורת המדינה מחייבת שהדיון יתחיל מן המוסד ולא מן הטכנולוגיה. ביקורת המדינה איננה רק בחינה של ממשל, ואף לא מסגרת כללית של פיקוח מוסדי. היא מתודולוגיה ופרקטיקה מוסדית-ציבורית מובחנת, הכוללת בדיקה יוזמת, מקצועית ואובייקטיבית²¹. צוות הביקורת נדרש לאסוף ולנתח ראיות, לבסס ממצאים, להסיק מסקנות ולגבש המלצות. את כל אלה עליו להפוך לדוח ביקורת המשמש בעת ובעונה אחת כלי לשיפור פעולת הגוף המבוקר וכלי פרלמנטרי וציבורי לפיקוח ולבקרה²². מכאן שהשימוש בכלי AI נבחן בביקורת המדינה לא רק לפי איכות התובנה שהוא מפיק, אלא לפי יכולתו להשתלב בתהליך ביקורת המוביל לדוח מבוסס, מנומק ותקף.

ייחודיות זו נשענת גם על עקרונות היסוד של ביקורת המגזר הציבורי. עקרונות ISSAI מדגישים כי ביקורת ציבורית בנויה על השגה והערכה אובייקטיבית של ראיות, כדי לאפשר השוואה שלהן לנורמות, וכי על מנהל הביקורת להשיג ראיות מתאימות ומספיקות לצורך ביסוס ממצאים, מסקנות והמלצות²³. במילים אחרות, ייחודיותה של ביקורת המדינה היא בעת ובעונה אחת מוסדית, מקצועית ונורמטיבית. מעמד זה נגזר בראש ובראשונה מן התלות המובנית של הביקורת בנורמות. לא ניתן להצביע על ליקוי בפעילות גוף מבוקר בלי להסתמך על נורמה, וקביעת ממצאי ביקורת על יסודות ברורים של נורמות ביקורת חשובה לשם שמירה על

20 Jakob Mökander, *Auditing of AI: Legal, Ethical and Technical*, לעיל ה"ש 13; Genaro-Moya, López-Hernández & Godz, *Artificial Intelligence 20 Approaches*, 2 Digital Society 49 (2023); Jakob Mökander & Luciano Floridi, *Operationalising AI Governance Through Ethics-Based Auditing: An Industry Case Study*, 3 AI and Ethics 451 (2023).

21 משרד מבקר המדינה **המדריך לביקורת המדינה**, לעיל ה"ש 1, לתהליך הביקורת השלם, בעמ' 1 - 4.

22 שם, רכיבי דוח הביקורת: תמונת מצב העולה מן הביקורת, עיקרי המלצות הביקורת וסיכום התקציר, בעמ' 3 - 5.

23 שם, ראיות הביקורת: הערכת ראיות ביקורת, בעמ' 1; INTOSAI, *ISSAI 100*, לעיל ה"ש 11; INTOSAI, *ISSAI 300*, לעיל ה"ש 11.

24 משרד מבקר המדינה, שם, לעיל ה"ש 1, הנורמות המשמשות את הביקורת: מהן נורמות ביקורת?, בעמ' 1.

25 שם, גיבוש ממצא ביקורת: מהו ממצא ביקורת?, בעמ' 1 - 2; שם, גיבוש מסקנת ביקורת: מהי מסקנת הביקורת? וכן רכיבי מסקנת הביקורת, בעמ' 1 - 2; שם, גיבוש המלצת ביקורת: מהי המלצת ביקורת? וכן דגשים לגיבוש המלצה, בעמ' 1 - 2.

26 שם, גיבוש מסקנת ביקורת: תיאור התוצאות ודגשים לגיבוש מסקנת ביקורת, בעמ' 1 - 2.

שנגזרו ממנה²⁷; הצדקה ביקורתית עוסקת בשאלה אם ממצא, מסקנה או המלצה בוססו כדין על נורמה וראיות²⁸.

ההבחנה בין אחריותיות לבין הצדקה ביקורתית היא ההבחנה בין הסבר על התהליך לבין ביסוס של הטענה. אחריותיות עונה על השאלה כיצד נעשה שימוש ב-AI: אילו נתונים הוזנו, אילו מודלים הופעלו, אילו בדיקות בוצעו ומי אישר את התהליך. לעומת זאת, הצדקה ביקורתית עונה על השאלה אם אפשר להוכיח את הממצא או המסקנה על בסיס נורמה וראיות. אפשר, למשל, להציג תיעוד מלא ושקוף של שימוש ב-AI לניתוח נתונים תקציביים (אחריותיות), אך עדיין אי אפשר לבסס ממצא בדוח ביקורת אם לא הוצגו ראיות ישירות, מאומתות ורלוונטיות לליקוי הנטען (הצדקה ביקורתית).

לכן ייחודיות ביקורת המדינה היא נקודת מוצא מתודולוגית מכרעת. כל דיון על שימוש בכלי AI חייב להיבחן לאור אופייה של הביקורת כמתודולוגיה ופרקטיקה מוסדית-ציבורית של השוואה לנורמות, איסוף ראיות, גיבוש ממצאים, ניסוח מסקנות והמלצות והצגתן בדוח שנועד לעמוד בבחינה מקצועית, ציבורית ופרלמנטרית. לאור זאת אפשר לברר מדוע גם מסגרות ממשל AI המפותחות ביותר עדיין אינן מספקות, כשהן לעצמן, מסגרת תקפה לשימוש בכלי AI במסגרת ביקורת מדינה.

ממשל AI ו-AI אמין/אחראי: בסיס נורמטיבי רחב, אך לא מסגרת מספקת לביקורת המדינה

לנוכח הדרישות המחמירות של תקני הביקורת כאמור, אפשר לבחון את כלי ה-AI מנקודת מבט פוזיטיבית כדלקמן:

א. ממשל AI הוא מסגרת רחבה לארגון עקרונות, תפקידים, סמכויות, מנגנוני בקרה, חובות תיעוד וניהול סיכונים לאורך מחזור השימוש של מערכת AI במוסדות ציבוריים²⁹. הוא עוסק, אפוא, בהבניה המוסדית של השימוש במערכת: מי אחראי, אילו סיכונים יש לנהל, כיצד מתעדים, מהן הבקורות הנדרשות ואיך מפקחים. במובן זה, ממשל הוא תנאי הכרחי לכל שימוש מוסדי ב-AI, אך הוא עדיין מסגרת-על ולא לוגיקת בחינה של השימוש בתוצרי AI.

ב. AI אמין ו-AI אחראי מתפקדים בתוך מסגרת זו כמצפן נורמטיבי-ערכי. הם מסמנים את התכונות ואת התנאים הרצויים שמערכת AI אמורה לקדם או לעמוד בהם: בטיחות, אמינות, הוגנות, שקיפות, פיקוח אנושי, כיבוד זכויות, אחריות והימנעות מפגיעה³⁰. בכך הם משרטטים את השאלה מה ראוי לצפות ממערכת AI או מן הארגון המשתמש בה. ואולם מצפן נורמטיבי, חשוב ככל שיהיה, עדיין איננו שקול למבחן ביקורתי. הוא מגדיר מה ראוי לבחון ומהם הערכים הרצויים, אך לא כיצד יש לבדוק, לאמת ולבסס אותם במסגרת ביקורתית.

ג. האחריותיות האלגוריתמית מוסיפה רובד מובחן נוסף. עניינה איננו רק בערכים הרצויים או במבנה הניהולי, אלא ביכולת לדרוש דין וחשבון, הסבר והצדקה בנוגע להחלטות, לתהליכים ולתוצרים שנגזרו מן השימוש בכלי AI³¹. תרומתה חשובה במיוחד משום שהיא מקרבת את השיח בעניין AI לשפה של הצדקה והסבריות. ואף על פי כן, גם אחריותיות אינה מספקת לבדה מסגרת ביקורתית שלמה. כשלעצמה היא עדיין אינה מבהירה מה בדיוק נבדק, אילו ראיות נדרשות וכיצד אפשר

27 ראו הרחבה לעיל בה"ש 4.

Mark Bovens, *Analysing and Assessing Accountability: A Conceptual Framework*, 13 European Law Journal 447 (2007); Madalina 28 Busuioc, *Accountable Artificial Intelligence: Holding Algorithms to Account*, 81 Public Administration Rev. 825 (2021); David Horneber & Sven Laumer, *Algorithmic Accountability*, 65 Bus. & Information Systems Engineering 723 (2023).

29 ראו לעיל ה"ש 9.

(High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI* (European Commission 2019 30 <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>; High-Level Expert Group on Artificial Intelligence, *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-Assessment* (European Commission 2020) Jobin, Ienca & <https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment> Floridi et al., *AI4People*; לעיל ה"ש 7; Vayena, *The Global Landscape* לעיל ה"ש 7.

31 Busuioc, *Accountable*, לעיל ה"ש 28; Horneber & Laumer, *Algorithmic*, לעיל ה"ש 28; Bovens, *Analysing*, לעיל ה"ש 28.

ביקורת המדינה: ממשל מצביע על הערך או על החובה; מתרגמת אותם לשאלה ביקורתית בת-בדיקה.

מגבלה זו מתחדדת עוד יותר כשמדובר בדרישות בנוגע לתיעוד, לשקיפות ולפיקוח אנושי במסגרת שימוש בכלי AI. מסמכי ממשל שונים מייחסים להם תפקיד מרכזי, ובצדק; אולם מבחינת ביקורת המדינה, עצם קיומם של תיעוד, רשומות, הסברים או מנגנוני oversight³⁵ עדיין אינו מספיק לשם תיקוף תוצרי AI במסגרת שרשרת הביקורת. תיעוד עשוי להיות תנאי חשוב לביקורת; הוא אינו סוף הביקורת. שקיפות עשויה לשפר את הנראות של ההליך; כשהיא לעצמה היא אינה מוכיחה את תוקפו. פיקוח אנושי מחויב להיות דרישה מוסדית חיונית; אך גם הוא טעון תרגום לשאלה ביקורתית: מה בדיוק נבדק בפיקוח, כיצד נבדק, ומהי הראיה לכך שהפיקוח היה אפקטיבי ולא פורמלי בלבד³⁶. לכן גם רשומות תיעוד לשלבי העבודה עם כלי AI, תיעוד לנתוני הקלט/הפלט, רשומות סיכון או דוחות שקיפות אינם ראיה מספקת מעצם קיומם³⁷; הם לכל היותר חומרי מוצא שיש לבחון, לאמת ולשלב במערך ראיות³⁸.

מכאן נגזר הצורך ב"שכבת תרגום". שכבה זו היא מסגרת מושגית-נורמטיבית המתווכת בין עקרונות השיח על AI לבין דרישות הראיה, הממצא וההצדקה של ביקורת המדינה, באמצעות המרת העקרונות והתאמתם לעולם המושגי של ביקורת המדינה³⁹. שכבת התרגום אינה מוסיפה עקרונות חדשים ואינה מחליפה את לוגיקת הביקורת; היא גם איננה מודל טכנולוגי או מתודולוגי עבודה מפורטת. תפקידה מצומצם ומדויק יותר: לשמש שכבת ביניים מושגית-

לבסס ממצא. במילים אחרות, אחריותיות עדיין איננה שקולה לביסוס ראיתי³².

מכאן נובעת הטענה המרכזית של סעיף זה: ממשל AI, AI, אמין/אחראי ואחריותיות הם שלושה רבדים חשובים אך שונים, ואף אחד מהם לבדו אינו מספק את מה שביקורת המדינה זקוקה לו. ממשל AI מספק מסגרת ארגונית; AI אמין/אחראי מספק מצפן נורמטיבי; אחריותיות מספקת דרישת דין וחשבון והצדקה. ואולם ביקורת המדינה זקוקה לצעד נוסף: תרגום של הרבדים הללו לקריטריוני ביקורת, להיבטי ראיות ולכללים להסקת מסקנות³³. לדוגמה, עיקרון של "פיקוח אנושי" במסגרת AI אמין קובע כי החלטות המבוססות על AI צריכות להיות נתונות לבקרה אנושית. ואולם בביקורת המדינה אין די בעצם קיומו של עיקרון זה. נדרש לתרגם אותו לקריטריוני ביקורת קונקרטיים: מי הגורם המאשר, אילו פעולות ביקורת עליו לבצע, כיצד מתועד האימות האנושי, ומה ייחשב ראיה מספקת לכך שהפיקוח לא היה פורמלי בלבד. רק לאחר תרגום קונקרטי כזה אפשר לבחון אם התקיים פיקוח אנושי אפקטיבי ולהסיק מסקנה ביקורתית מבוססת.

מגבלה זו מתבלטת בייחוד לאור ההבחנה בין עקרונות לקריטריוני ביקורת. עיקרון מצביע על ערך רצוי; קריטריון ביקורת מאפשר לבחון התאמה. גם כאשר הביקורת נשענת על עקרונות כלליים או על נורמות רצויות, יש לגבש קריטריונים רלוונטיים שיאפשרו לבדוק את מידת ההתאמה של הנושא הנבדק לעקרונות אלו³⁴. דרישה זו מלמדת באופן חד כי ממשל אינו לוגיקת העבודה וההוכחה של

32 משרד מבקר המדינה המדריך לביקורת המדינה, לעיל ה"ש 1, ראיות הביקורת: הערכת ראיות ביקורת, בעמ' 1 - 2; Busuioc, *Accountable*, לעיל ה"ש Horneber & Laumer, *Algorithmic*, לעיל ה"ש 28.

33 Mokander & Floridi, *Ethics-Based*, לעיל ה"ש 11; Mokander, *Auditing of AI*, לעיל ה"ש 20; Mokander & Floridi, *Operationalising*, לעיל ה"ש 20; INTOSAI, *ISSAI 100*, לעיל ה"ש 11.

34 משרד מבקר המדינה המדריך לביקורת המדינה, לעיל ה"ש 1, גיבוש ממצא ביקורת: סוגי ממצאי ביקורת, בעמ' 1 - 2; שם, הנורמות המשמשות את הביקורת: מהן נורמות ביקורת?, בעמ' 1 - 2.

35 מנגנוני oversight (פיקוח) הם מנגנונים של בקרה, בחינה ואתגור של השימוש במערכת AI. במאמר זה עצם קיומו של מנגנון oversight אינו מספיק כשלעצמו, אלא נדרש לבחון את תוכנו ואת מידת האפקטיביות שלו.

36 משרד מבקר המדינה המדריך לביקורת המדינה, לעיל ה"ש 1, ראיות הביקורת: הערכת ראיות ביקורת, בעמ' 1 - 2; Busuioc, *Accountable*, לעיל ה"ש 28; Horneber & Laumer, *Algorithmic*, לעיל ה"ש 28.

37 ראיה מספקת היא ראיה שנבחנה, אומתה והוצלבה באופן שמאפשר לבסס ממצא, מסקנה או המלצה במסגרת שרשרת הביקורת. לכן תוצר AI אינו יכול להיחשב ראיה מספקת גם אם יש תיעוד, דוח שקיפות או רשומות סיכון בנוגע אליו, אלא רק תוצר ראשוני שיכול להשתלב במערך שרשרת הביקורת לאחר שעבר בחינה ביקורתית.

Margaret Mitchell et al., *Model Cards for Model Reporting*, Proceedings of the Conference on Fairness, Accountability, and Transparency 38 220-229 (2019); Timnit Gebru et al., *Datasheets for Datasets*, Proceedings of the 5th Workshop on Fairness, Accountability, and Transparency in Machine Learning (2018); Khoa Lam et al., *A Framework for Assurance Audits of Algorithmic Systems*, ARXIV (January 2024), arXiv:2401.14908 [cs. CY].

Mittelstadt, *Principles Alone*, לעיל ה"ש 12; Morley, *From What to How*, לעיל ה"ש 12; Mokander & Floridi, *Ethics-Based*, לעיל ה"ש 11.

אפיסטמית הקובעת את תנאי המעבר התקפים בין שיח ה-AI לבין לוגיקת הביסוס וההוכחה של ביקורת המדינה⁴⁰. שכבת התרגום אינה מערכת כללים אופרטיבית מלאה, אלא מסגרת ביניים מושגית הקובעת כיצד אפשר לתרגם עקרונות AI לדרישות ביקורת תקפות.

שכבת התרגום פועלת בארבעה מישורים משלימים: היא ממירה עקרונות כלליים לקריטריוני ביקורת; היא תוחמת בין חומר גלם או תוצר אנליטי לבין ראייה תקפה, ומונעת קפיצה מ-AI לממצא, הפוגעת בתקפות הממצא; היא מעגנת את השימוש ב-AI בדרישות של רלוונטיות, מהימנות, מספיקות ויכולת אימות; והיא משלבת את השימוש ב-AI בדרישות המקצועיות של דוח הביקורת, ההצדקה הציבורית ואחריות מנהל הביקורת לדוח הסופי⁴¹.

רק במסגרת תרגום כזו אפשר יהיה לשלב שימוש בכלי AI בתהליך ביקורת המדינה באופן שאינו עוקף את השרשרת הפורמלית של נורמה - ראייה - ממצא - מסקנה - המלצה, אלא נטמע בה ומסייע לכל שלבי הביקורת באופן תקף ומבוקר⁴².

אחריות אלגוריתמית בהקשר של עבודת הביקורת

אחריות אלגוריתמית עניינה הוא חובת דין וחשבון, הסבר והצדקה בנוגע לאופן השימוש בכלי, לתהליך שהשימוש בו יצר ולתוצרים שנגזרו ממנו⁴³. אלא שבביקורת המדינה לחובה זו מוקנית משמעות מוסדית מחמירה יותר. עבודת הביקורת מתכנסת לדוח פומבי המשמש הן כלי עבודה לשיפור פעולת הגוף המבוקר הן כלי לפיקוח ציבורי ופרלמנטרי. לכן השאלה איננה רק אם אפשר להסביר את פעולת הכלי, אלא אם אפשר להצדיק את מקומו של השימוש בו בתוך תהליך ביקורת החשוף לבחינה חיצונית⁴⁴.

בהתאם לכך, אחריות אלגוריתמית רלוונטית לביקורת המדינה רק כאשר היא מתורגמת ליכולת לשחזר ולהצדיק את השתלבות הכלי בשרשרת עבודת הביקורת: מה היה תפקידו של תוצר ה-AI, כיצד נבחן, מה אומת מתוכו, מה נדחה, וכיצד שולב, אם בכלל, במהלך עבודת הביקורת. עם זאת, אין לזהות בין אחריות לבין ביסוס ראייתי. אפשר שתהליך השימוש בכלי יהיה מתועד היטב; אפשר שיוצגו רישומי סיכון, תיעוד נתונים או מנגנוני פיקוח אנושי; אך כל אלה, חשובים ככל שיהיו, עדיין אינם מקימים כשלעצמם ראיות ביקורת מספיקות ומתאימות לממצא, למסקנה או להמלצה⁴⁵. תיעוד של אופן הפקת התוצר איננו זהה לראיה המעידה על המצב העובדתי בגוף המבוקר, ושקיפות בנוגע להליך העבודה של הכלי אינה מוכיחה כשלעצמה כי התקיים פער מן הנורמה. לפיכך אחריות אלגוריתמית היא תנאי עזר חיוני לביקורת, אך לא תחליף לראיות הביקורת. ערכה הוא שהיא מונעת שימוש ב-AI כקופסה שחורה מבחינת צוות הביקורת, והיא מחייבת שאפשר יהיה להצדיק את השימוש שנעשה בו גם לאחר מעשה. דווקא משום כך היא חייבת להיקשר ללוגיקה הפורמלית של נורמה - ראייה - ממצא - מסקנה - המלצה, ולא להישאר ברמת שיח כללית על שימוש אחראי בטכנולוגיה⁴⁶.

אחריות אלגוריתמית עשויה לתמוך בבחינה הביקורתית, אך היא איננה חוליה חלופית בשרשרת נורמה - ראייה - ממצא - מסקנה - המלצה, ואינה מקנה מעמד ראייתי לתוצר ה-AI.

נורמות ביקורת וראיות ביקורת כתנאי סף לשימוש תקן ב-AI

הלב האפיסטמי של טענתנו בדבר הצורך ב"שכבת תרגום" מצוי במפגש שבין נורמות לראיות. ביקורת המדינה אינה יכולה להצביע על ליקוי בלי להסתמך על נורמה,

Mokander, *Auditing of AI* 40, לעיל ה"ש 20.

41 משרד מבקר המדינה **המדריך לביקורת המדינה**, ראיות הביקורת: "הערכת ראיות ביקורת", בעמ' 1 - 2; שם, רכיבי דוח הביקורת: תמונת מצב העולה מן הביקורת, עיקרי המלצות הביקורת וסיכום התקציר, בעמ' 3 - 5; שם, תהליך הביקורת השלם, בעמ' 3 - 4.

42 INTOSAI, *ISSAI 100*, לעיל ה"ש 11; INTOSAI, *ISSAI 300*, לעיל ה"ש 11; INTOSAI, *ISSAI 400*, לעיל ה"ש 11; Genaro-Moya, López-Hernández & Godz, *Artificial Intelligence*, לעיל ה"ש 13.

43 Busuioc, *Accountable*, לעיל ה"ש 28; Horneber & Laumer, *Algorithmic*, לעיל ה"ש 28; Bovens, *Analysing*, לעיל ה"ש 28.

44 Genaro-Moya, López-Hernández & Godz, *Artificial Intelligence*, לעיל ה"ש 13.

45 Mittelstadt, *Principles Alone*, לעיל ה"ש 12; Morley, *From What to How*, לעיל ה"ש 12; Busuioc, *Accountable*, לעיל ה"ש 28; Horneber & Laumer, *Algorithmic*, לעיל ה"ש 28; Bovens, *Analysing*, לעיל ה"ש 28.

46 INTOSAI, *ISSAI 100*, לעיל ה"ש 11; INTOSAI, *ISSAI 300*, לעיל ה"ש 11; INTOSAI, *ISSAI 400*, לעיל ה"ש 11; Genaro-Moya, López-Hernández & Godz, *Artificial Intelligence*, לעיל ה"ש 13.

מעמדו הראשוני הוא של חומר ביניים אנליטי. כדי שמידע שנעזר ב-AI יתרום לביסוס ממצא, עליו לעבור עיבוד ביקורתי: בירור מקורו, בחינת הקשרו, אימות, הצלבה עם מקורות נוספים, מיקומו ביחס לנומרה ושילובו במערך ראיות מספיק ומתאים. רק לאחר תהליך כזה אפשר לבחון אם המידע שנעזר ב-AI אכן עומד בתנאי הקבלה של ראיית הביקורת.

המעבר התקף הוא אפוא זה: תוצר AI ← עיבוד ביקורתי ← ראיות ← ממצא. אפשר לתאר זאת גם כמבחן כניסה כפול. ראשית, על התוצר שנעזר ב-AI לעמוד במבחן אפיסטמי: להיות ניתן לאימות, להצלבה ולבירור מקורו והקשרו. שנית, עליו לעמוד במבחן ביקורתי: להיות ניתן למיקום ביחס לנומרה הרלוונטית ולהשתלב במערך ראיות שממנו אפשר לעבור כדי לממצא. כל עוד אחד משני המבחנים הללו אינו מתקיים, התוצר נותר במעמדו האנליטי בלבד ואינו נכנס למרחב הראיות.

הבחנה זו מבהירה גם את מעמדם של כלי תיעוד ובקרה⁴⁹ כגון model cards, datasheets, רשומות שימוש, דוחות שקיפות ומנגנוני oversight⁵⁰. כל אלה עשויים להיות חשובים להבנת מגבלות הכלי, לתיעוד תהליך העבודה ולחזיק יכולת השחזור והבקרה. לעיתים הם אף יהיו חיוניים כדי להבין אם ראוי להסתמך על תוצר מסוים לצורך המשך בדיקה. אולם הם ודאי עדיין אינם שקולים לראיות ביקורת על הנושא הנבדק עצמו. הם עשויים ללמד כיצד הופק תוצר בעזרת כלי AI; אבל אין הם מוכיחים כשלעצמם כי הגוף המבוקר פעל בניגוד לנומרה, כי נגרם פער, או כי מוצדק לגזור ממנו מסקנה והמלצה. לכן תנאי הסף לשימוש תקף ב-AI אינם נקבעים לפי איכות התיעוד שלו ואף לא לפי רמת ההסבריות שלו בלבד, אלא לפי השאלה אם לאחר עיבוד ביקורתי אפשר לשלב את תוצריו במערך ראיות רלוונטי, תקף, מהימן ומספיק.

והיא אינה יכולה לבסס ממצא בלי להסתמך על ראיות. הנורמות המשמשות את הביקורת כוללות הוראות מחייבות הקבועות בדיון, עקרונות כלליים של מינהל תקין ולעיתים גם נורמות רצויות שמפותחות במסגרת הביקורת על בסיס נורמות קיימות והתאמתן לנסיבות הנבדקות.

בצד הנומרה ניצבות ראיות הביקורת. לפי הנחיות משרד מבקר המדינה ולפי תקני ISSAI, ראיות הביקורת צריכות להיות מתאימות ומספיקות. התאמתן נבחנת לפי רלוונטיות, תקפות ומהימנות; ומספיקות היא עיקרון כמותי הקובע כי על הצוות לאסוף די ראיות לתמיכה בממצא⁴⁷. לכן לא די בכך שמידע מסוים "נראה נכון", "מתקבל על הדעת" או "מנוסח היטב". עליו לעמוד במבחני הראיה של הביקורת. במובן זה, הדרישה הראייתית של ביקורת המדינה מחמירה דווקא מפני שהיא איננה הליך שיפוטי ואינה כפופה לדיני הראיות במובנם הפורמלי. מאחר שאין מדובר במערכת קבילות שיפוטית אלא במנגנון מקצועי של ביסוס והנמקה, מוטלת על צוות הביקורת אחריות מוגברת לוודא כי ממצאיו נשענים על חומר המספיק מבחינה כמותית ומתאים מבחינה איכותית. הדרישה לראיות רלוונטיות, תקפות, מהימנות ומספיקות איננה תחליף לדיני הראיות, אלא תנאי תוקף פנימי של הביקורת עצמה. מבחינה זו, שימוש ב-AI שאינו ניתן לאימות, שאינו ניתן להצלבה או שאינו מאפשר להבין מה מקורו ומה מעמדו איננו רק שימוש טכנולוגי בעייתי; הוא שימוש שאינו יכול להיכנס מלכתחילה למרחב ההוכחה של הביקורת⁴⁸.

מכאן נובעת הטענה המרכזית של פרק זה: **מידע שהופק בעזרת AI אינו נכנס מאליו למרחב הראיות של הביקורת. מעמדו הראשוני הוא של חומר אנליטי הטעון עיבוד ביקורתי, אימות, הצלבה ומיקום ביחס לנומרה, ורק לאחר מכן אפשר לבחון אם וכיצד הוא עשוי להשתלב במערך הראיות.** גם כאשר התוצר מועיל, מהיר או מדויק לכאורה,

47 משרד מבקר המדינה **המדריך לביקורת המדינה**, ראיות הביקורת: הערכת ראיות ביקורת, בעמ' 1; INTOSAI, ISSAI 100, לעיל ה"ש 11; INTOSAI, ISSAI 300, לעיל ה"ש 11.

48 Khoo, A Framework, Mitchell, Model Cards, לעיל ה"ש 38; Timnit, Datasheets, לעיל ה"ש 38.

49 תיעוד ובקרה אינם עקרונות עצמאיים בלבד במסגרת ממשל AI, אלא רכיבים הנגזרים מעקרונות רחבים יותר, כגון שקיפות, עקיבות, אחריותיות ופיקוח אנושי. עם זאת, בהקשר של ביקורת המדינה, גם כאשר מתקיימים מנגנוני תיעוד ובקרה, אין די בכך כדי לבסס ראיית ביקורת או להצדיק ממצא.

50 model cards הם מסמכי תיעוד שיטתיים המתארים מאפיינים מרכזיים של מודל AI, כגון ייעוד, מגבלות, נתוני אימון, ביצועים, שימושים מותרים וסיכונים. במאמר זה הם נחשבים כלי תיעוד ובקרה, לא ראיית ביקורת כשלעצמם.

datasheets הם מסמכי תיעוד שיטתיים לנתונים או למאגרי נתונים, הכוללים מידע על מקורם, על אופן איסופם, על מבנם, על מגבלותיהם ועל הקשרי השימוש בהם. גם הם עשויים לתרום לעקיבות ולבקרה, אך אינם שקולים לראיות ביקורת.

רשומות שימוש הן תיעוד שיטתי של אופן השימוש בפועל במערכת AI, ובכלל זה נתוני קלט ופלט, מועדי שימוש, זהות המשתמשים, שלבי העבודה והפעולות שבוצעו. במאמר זה רשומות שימוש נתפסות ככלי לעקיבות, לתיעוד ולבקרה, אך אינן שקולות כשלעצמן לראיית ביקורת.

דוחות שקיפות הם מסמך רחב, בדרך כלל ברמת ארגון או מסגרת השימוש. הם מתארים איך נעשה שימוש במערכת, אילו כללי בקרה חלים עליה, אילו סיכונים זוהו, איך מטפלים בתלונות, בחריגות או בכשלים ואילו צעדי פיקוח, תיעוד ומדיניות קיימים.

תחום התחולה של השימוש בכלי AI בתהליך העבודה של ביקורת המדינה

תהליך הביקורת השלם כולל חמישה חלקים: תכנון, ביצוע - "עבודת השדה", גיבוש הדוח וכתביבתו, המצאת הדוח ופרסומו, ומעקב אחר תיקון הליקויים⁵¹. תחום התחולה של השימוש ב-AI צריך להיקבע לאורך מחזור זה לא כרשימת שימושים אפשריים, אלא לפי היחס שבין כל שלב לבין שרשרת ההצדקה הביקורתית. ככל שהשלב קרוב יותר לבניית בסיס הידע, למיפוי החומר, לארגון ולזיהוי כיווני בדיקה, כך עשויה תרומתו האנליטית של AI להיות רחבה יותר. ככל שהשלב קרוב יותר לגיבוש ממצא, מסקנה או המלצה, כך השימוש התקף בכלי נעשה צר יותר, ותלוי יותר באימות, בהצלבה, בתייעוד ובשיקול הדעת המקצועי של מנהל הביקורת.

הבחנה זו מלמדת כי תחום התחולה של AI אינו נחתך לפי קו טכני של "מותר" ו"אסור", אלא לפי דרגת הקרבה של הפעולה לליבת ההכרעה הביקורתית. במקום שבו העבודה מכוונת להרחבת שדה הבדיקה, למיון החומר או לזיהוי כיווני עיון, אפשר להכיר בתרומה אנליטית רחבה יחסית של הכלי. לעומת זאת, במקום שבו העבודה מכוונת להכריע מהו המצב בפועל ביחס לנומרה, האם הוכח פער, מה משמעותו ומהי ההמלצה הנגזרת ממנו, השימוש התקף בכלי צריך להיעשות משני, מסייע ומגביל. זוהי לא רק הבחנה מעשית, אלא גם הבחנה אפיסטמית: ככל שמתקדמים מן החיפוש והעיבוד אל ההכרעה וההצדקה, כך עולה משקלן של ראיות הביקורת ושל הנמקה האנושית, ופוחתת היכולת להישען על תוצר AI כבסיס ישיר להמשך הדרך. מכאן שההבחנה בין שלבי התהליך איננה הבחנה מתודולוגית-תפעולית בלבד. היא קובעת את היחס בין תרומה אנליטית אפשרית לבין מעמד ביקורתי תקף: בשלבים המקדמיים AI עשוי להרחיב את יכולת העיבוד; בשלבי ההכרעה הוא אינו יכול לשמש אלא ככלי מסייע, ולעולם לא מקור עצמאי לקביעת ממצא, מסקנה או המלצה⁵². גם כאשר AI מסייע בזיהוי חלופות, מוקדי סיכון או כיווני בדיקה, ההכרעה מהן שאלות הביקורת ומהן הנורמות הרלוונטיות נותרת הכרעה מקצועית של מנהל הביקורת.

בשלב התכנון - לרבות הלימוד הראשוני, הסקר, הגדרת יעדי הביקורת ושאלותיה, מיפוי הנורמות ובניית בסיס הידע - AI עשוי לתרום תרומה ניכרת לארגון היקף חומר גדול, לאיתור נושאים חוזרים, לזיהוי מוקדי סיכון אפשריים ולהפקת תמונת רקע ראשונית. ואולם גם כאן אין הוא רשאי

להגדיר את שאלות הביקורת, את היקפה, את הנורמות או את הקריטריונים שישמשו בה. הכרעות אלה הן לב תכנון הביקורת, ולכן הן נותרות באחריותו המקצועית של מנהל הביקורת. בשלב עבודת השדה - שבו על הצוות לאסוף ולנתח ראיות ולהשיג ראיות מספיקות והולמות כדי לבסס ממצאים, מסקנות והמלצות - AI עשוי לשמש כלי עזר אנליטי רב ערך: לסיווג מסמכים, לאיתור חריגות או דפוסים חוזרים הדורשים בדיקה, לתמצות חומרי רקע או לניתוח ראשוני של נתונים. עם זאת, גם אם AI מסייע לזהות כיוון בדיקה חשוב וגם אם הוא מסייע לארגון תשתית חומר, אין בכך כדי לפטור מן הצורך להעריך את איכותה הראייתית ולהפעיל את שיקול הדעת. בשלב זה לכל היותר אפשר לומר שהכלי מרחיב את יכולת העיבוד של מנהל הביקורת, אך הוא אינו מקצר את מסלול ההוכחה.

בשלב גיבוש הדוח וכתביבתו תחום התחולה מצטמצם עוד יותר. אפשר להיעזר ב-AI לצורכי ארגון חומר, בדיקת עקיבות פנימית, זיהוי כפילויות וסתירות או ניסוח ראשוני זהיר של טקסט, אולם הממצאים, המסקנות וההמלצות עצמם חייבים להיות פרי של הנמקה ביקורתית המבוססת על ראיות הביקורת. שלב כתיבת הדוח אינו רק שלב ניסוחי; הוא שלב נוסף של הצדקה.

גם בשלבי המצאת הדוח, פרסומו, ריכוז הליקויים וההמלצות, המעקב אחר תיקון הליקויים והפקת הלקחים, ייתכן תפקיד תומך ל-AI בארגון מידע, באיתור תבניות חוזרות ובסיוע במעקב. אלא שבשלבים אלה כבר ברור כי עצם השימוש בכלי אינו מוסיף תוקף מהותי לממצא, למסקנה או להמלצה. לכן דווקא בשלבים אלה יש מקום רב יותר לשימוש ב-AI ככלי מסייע לניהול המידע, התהליך והמשך המעקב.

מכאן נובעת המסקנה כי השימוש ב-AI בביקורת המדינה אפשרי, ולעיתים אף מועיל מאוד, אך הוא חייב להיות תחום על פי לוגיקת הביקורת ולא על פי פוטנציאל הטכנולוגיה בלבד. AI יכול לתרום ללמידה, לארגון, לזיהוי, למיון ולתמצות; הוא יכול לתמוך בתייעוד, בשחזור ובבדיקה; והוא יכול לסייע לצוות להתמודד עם מורכבות ועם היקף חומר גדול. אך ככל שמתקרבים לליבת ההכרעה הביקורתית, כך מתברר מחדש שהשאלה המכריעה איננה מה הכלי מסוגל להפיק, אלא מה אפשר לבסס באמצעותו במסגרת ביקורת מבוססת ראיות. בכך מושלמת שכבת התרגום שהמאמר מבקש להציע: לא מיפוי כללי של אפשרויות שימוש ב-AI, אלא תחימה מושגית-נורמטיבית של התנאים שבהם AI עשוי לתרום תרומה אנליטית תקפה, ושל הגבולות

51 משרד מבקר המדינה המדריך לביקורת המדינה, לעיל ה"ש 1, תהליך הביקורת השלם, בעמ' 1 - 3.

52 שם, ראיות הביקורת: הערכת ראיות ביקורת, בעמ' 1 - 2.



לאחר עיבוד כזה יכול מידע שנעזר ב-AI להשתלב בראיות הביקורת, שעל בסיסן אפשר לגבש ממצא. ייחוס מעמד ראייתי ישיר לתוצר AI אינו עולה בקנה אחד עם מבנה ההיסק של הביקורת.⁵³

הבחנה זו נובעת מן ההגדרות הפורמליות של עבודת הביקורת. ראיית ביקורת היא מידע המשמש את צוות הביקורת כדי לתקף ולבסס את ממצאיו ולהסיק מסקנות ברמת ודאות מספקת; ממצא ביקורת הוא מידע עובדתי רלוונטי הנובע מהשוואה בין הנורמה למצב בפועל, כפי שהוא עולה מראיות הביקורת. **מכאן שהמעבר הישיר AI ← ממצא איננו רק קיצור דרך פסול, אלא הוא שלילה של מבנה ההיסק הביקורתי.** המסקנה המתבקשת מכך היא שהמעבר התקף הוא: תוצר AI ← עיבוד ביקורתי ← ראיות ← ממצא.

מן ההבחנה הראשונה נגזרת הבחנה שנייה: תובנה המופקת בעזרת AI איננה ממצא. תובנה, גם כאשר היא חדה, מועילה או משכנעת, היא לכל היותר איתות אנליטי ראשוני, כיוון בדיקה או השערת עבודה. ממצא, לעומת זאת, הוא קביעה עובדתית רלוונטית המעוגנת בראיות ומנוסחת ביחס לנורמה. לפיכך גם תוצר AI המפיק תובנה בעלת ערך אינו מקבל מכוח זה מעמד של ממצא. בהקשר הביקורתי הסיכון איננו רק טעות עובדתית, אלא המרת שכנוע אנליטי או לשוני למעמד עובדתי שלא נבנה על בסיס ראיות.

שמעבר להם אין הוא יכול לשמש בסיס לממצא, למסקנה או להמלצה.⁵³

הבחנות מושגיות: AI אינו ראייה; תובנה אינה ממצא

עד כה פרקי המאמר ביססו את התנאים הנורמטיביים, הראייתיים והמתודולוגיים להשתלבות תקיפה של AI בביקורת המדינה. על בסיס זה, פרק זה איננו מוסיף תנאי סף חדשים, אלא מרכז, מחדד ותוחם את ההבחנות המושגיות הקובעות מה אין להסיק מתוצר AI במבנה ההיסק של הביקורת: תוצר AI אינו ראייה כשלעצמו, ותובנה המופקת על ידי AI אינה ממצא. תפקידן של הבחנות אלה הוא לתחום במדויק את מקומו של תוצר AI בהליך הביקורת, ולמנוע ייחוס שגוי של מעמד ראייתי או ביקורתי לתוצר שאיננו נושא אותו.⁵⁴

ההבחנה הראשונה, והקובעת מכולן, היא **שתוצר AI כשלעצמו איננו ראיית ביקורת**. מעמדו הראשוני הוא של חומר אנליטי ראשוני: הוא עשוי לסייע בארגון מידע, בסיווג, באיתור כיווני בדיקה, בזיהוי חריגות לכאורה ובהפקת תובנה ראשונית; אולם אין בו, כשלעצמו, משום ראייה. תוצר AI עשוי לשמש חומר אנליטי טעון עיבוד ביקורתי, הכולל בדיקה, אימות, הצלבה ומיקום ביחס לנורמה. רק

53 Mokander, *Auditing of AI*, לעיל ה"ש 20; Genaro-Moya, López-Hernández & Godz, *Artificial Intelligence*, לעיל ה"ש 13.

54 INTOSAI, *ISSAI 100*, לעיל ה"ש 11; INTOSAI, *ISSAI 300*, לעיל ה"ש 11; משרד מבקר המדינה **המדריך לביקורת המדינה**, לעיל ה"ש 1, ראיות הביקורת: מהי ראיית ביקורת? וכן הערכת ראיות ביקורת, בעמ' 1-2; שם, גיבוש ממצא ביקורת: מהו ממצא ביקורת, בעמ' 1-2; שם, תהליך הביקורת השלם, בעמ' 2-3.

55 משרד מבקר המדינה **המדריך לביקורת המדינה**, לעיל ה"ש 54.

סיכונים, מגבלות אפיסטמיות ושרשרת הראיות הביקורתית

מגבלותיהם של כלי AI, ובייחוד של מערכות גנרטיביות, אינן רק "תופעת לוואי" של איכות טכנולוגית; הן נוגעות ישירות לאפשרות לבסס ממצא תקף. כך למשל, כחלק מהתכונות המבניות של כלי AI התוצר שלהם נוצר בדרך של חיזוי הסתברותי של רצפי מידע, ולא על בסיס הבנה, ידיעה או אימות עובדתי עצמאי. לכן גם כאשר יצירת התוצר מסתמכת רק על מקורות מידע שהוגדרו במפורש, ללא הוספת מידע או השלמת מידע על ידי כלי ה-AI, ייתכן תוצר שנראה תקין ומתאים למקורות המידע אך בפועל משקף לוגיקה הסתברותית שונה מזאת שבמקורות המידע⁵⁷. הספרות על אי-ודאות ועל הזיה מלמדת כי אי-ודאות איננה אנומליה שולית, אלא תכונה מבנית של מערכות למידה, וכי הזיה היא דפוס כשל המחייב מנגנוני אימות, ריסון והימנעות מהסקה ישירה. בהקשר של ביקורת המדינה משמעות הדבר איננה רק שעל מנהל הביקורת "להיזהר"; משמעותו היא שפלט AI בלתי מאומת אינו יכול לשמש כשלעצמו ראיית ביקורת, וכי רק לאחר בדיקה, אימות והצלבה אפשר לבחון אם מידע שהופק בעזרת AI עשוי להשתלב בראיות הביקורת. כאשר דוח הביקורת מבקש לייחס לגוף מבוקר פער, ליקוי, תוצאה או סיכון, אין די בתוכן משכנע או בניסוח בטוח שהופק על ידי AI, אלא נדרש בסיס המאפשר הצדקה ראיתית⁵⁸.

אלא שהקושי אינו נובע רק מסיכון של הזיה או אי-ודאות. גם כאשר הפלט נראה תקין, השאלה הביקורתית איננה רק אם אפשר להבין פלט יחיד, אלא אם אפשר לשחזר את תנאי הפקתו, בדיקתו והשימוש שנעשה בו: מהם מקורות המידע, איזו גרסה שימשה, אילו בדיקות בוצעו, אילו מגבלות היו ידועות, מי בחן, מי אישר, ואילו אירועי "override"⁵⁹, תיקון או דחייה התרחשו. לכן עקיבות ותיעוד אינם רק דרישות של ממשל AI תקין; הם תנאי עזר הכרחיים לביחנית האפשרות הראיתית.

עקרונות נוספים בשיח הממשל AI, כגון שקיפות, עקיבות, תיעוד ורכיבי בקרה (כגון model cards, datasheets, ורשומות שימוש), עשויים לסייע בהבנת מקורות הנתונים, תנאי ההפעלה, מגבלות המערכת ואופן השימוש בה, ובכך לתרום לביחנית ביקורתית מושכלת יותר. אולם ערכם הוא ערך מסייע בלבד: הם מאפשרים לבחון את דרך הפקת התוצר ואת תנאי השימוש בו, אך ברור שאינם מוכיחים כשלעצמם מה היה המצב בפועל בגוף המבוקר, אם התקיים פער מן הנורמה, או אם נאספו די ראיות רלוונטיות, מהימנות ומספיקות לביסוס ממצא.

מכאן מתחדדת גם ההבחנה בין שיח הממשל AI לבין לוגיקת הביקורת. עקרונות כגון שקיפות, פיקוח אנושי, אחריותיות, ניהול סיכונים ותיעוד מגדירים ציפיות נורמטיביות כלליות, אך אינם נטמעים מעצמם בשרשרת ההיסק של ביקורת המדינה, וודאי שהתקיימותם לא מביאה מכוח עצמם לביסוס ראיה, ממצא או מסקנה. כדי שאפשר יהיה להתאימם באופן תקף לדרישות לוגיקת הביסוס וההוכחה של ביקורת המדינה, נדרשת שכבת התרגום שהמאמר מציע: מסגרת מושגית-נורמטיבית המתווכת בין עקרונות השיח על AI לבין דרישות הראיה, הממצא וההצדקה של ביקורת המדינה. שכבה זו אינה מוסיפה עקרונות חדשים ואינה מחליפה את לוגיקת הביקורת; תפקידה לקבוע את תנאי המעבר התקפים בין שיח ה-AI לבין לוגיקת הביסוס וההוכחה של ביקורת המדינה. במסגרת זו היא פועלת בארבע דרכים משלימות: היא ממירה עקרונות כלליים לקריטריוני ביקורת אופרטיביים; היא תוחמת בין תוצר אנליטי לבין ראיה תקפה ומונעת קפיצה מ-AI לממצא; היא מעגנת את השימוש ב-AI בדרישות של רלוונטיות, מהימנות, מספיקות ויכולת אימות; והיא משלבת את השימוש בו בדרישות המקצועיות של דוח הביקורת, ההצדקה הציבורית ואחריותיות מנהל הביקורת לדוח הסופי. במובן זה, ההבחנות המושגיות אינן רק ניסוחים שליליים; הן תנאי הסף החיוניים לשימוש תקף ב-AI בביקורת המדינה⁶⁰.

56 INTOSAI, ISSAI 100, לעיל ה"ש 11; INTOSAI, ISSAI 300, לעיל ה"ש 11; Mittelstadt, *Principles Alone*, לעיל ה"ש 12; Morley, *From What to How*, לעיל ה"ש 12; Mökander, *Auditing of AI*, לעיל ה"ש 20; Mökander, *Operationalising AI*, לעיל ה"ש 20.

57 אפשר לדמות זאת לדף צביעה שיש בו קווי מתאר לציר מורכב מאוד. גם בהינתן כללים מנחים בנוגע להתאמות צבעים וגוונים, יש אפשרויות רבות לאופן צביעת הציר. במידה מסוימת כך גם בכלי AI שמייצרים תוצר באופן של חיזוי הסתברותי. תוצר AI עשוי להיראות קוהרנטי, הגיוני ואמין גם כאשר אין לו בסיס עובדתי מספיק או, לחלופין, כאשר אופן עיבוד המידע שונה מהלוגיקה האמיתית של המידע המקורי, ומכאן הצורך באימות והצלבה על ידי מנהל הביקורת ובהימנעות מהסקה ישירה.

58 Hüllermeier & Waegeman, *Aleatoric and Epistemic*, לעיל ה"ש 18; Huang et al., *A Survey on Hallucination*, לעיל ה"ש 18; Ziwei Ji et al., *Survey of Hallucination in Natural Language Generation*, 55 ACM Computing Surveys 248 (2023).

59 אירועי Override הם פעולה של ביטול, עקיפה או שינוי של החלטה, המלצה או פעולה שנעזרה ממערכת AI בידי גורם אנושי או מנגנון בקרה. אירועים אלה הם חלק מהפיקוח והבקרה על כלי AI, והידע לגביהם הוא חלק מדרשות התיעוד. במאמר זה המונח קשור לפיקוח ולבקרה, אך אינו מוכיח כשלעצמו אפקטיביות של פיקוח.

מכאן שהשימוש בכלי AI בביקורת המדינה מחייב לראות במגבלות אפיסטמיות ובדרישות עקיבות ותייעוד לא רשימת סיכונים טכנית, אלא חלק מתנאי היכולת לבחון אם אפשר לבסס ממצא תקף. כל עוד לא בוצע עיבוד ביקורתי המאפשר לבחון אם מידע שהופק בעזרת AI יכול להשתלב בראיות ביקורת רלוונטיות, תקפות, מהימנות ומספיקות, אין בסיס תקף לא לממצא, לא למסקנה ולא להמלצה. כך עובר מרכז הכובד מן השאלה מה רצוי או יעיל אל השאלה מה אפשר לבסס, לאמת ולהצדיק במסגרת דוח ביקורת מבוסס ראיות.

פיקוח אנושי ושיקול דעת מקצועי כחלק אינהרנטי מן ההליך

מכיוון שתוצר AI אינו ראייה, ומכיוון שהסבר, תיעוד או עקיבות עדיין אינם שקולים לביסוס ראיות מספיקות ומתאימות, מתחדדת השאלה מה מקומו של מנהל הביקורת עצמו בתוך ההליך. גם כאן אין די בעיקרון הכללי של *human in the loop* (אדם בתוך התהליך)⁶². עיקרון זה מצוין לכל היותר נוכחות של אדם במסגרת התהליך, אך אין בו כדי להבטיח פיקוח ביקורתי אפקטיבי. יש להבחין בין השתתפות פורמלית לבין הפעלה ממשית של שיקול דעת מקצועי, בדיקה, אתגור, הכרעה והנמקה, הנדרשים כתנאי להשתלבות תקפה של המידע שנוצר באמצעות AI בהליך הביקורת. פיקוח אנושי אפקטיבי איננו עצם נוכחותו של אדם בתהליך, אלא מנגנון של ספקנות ביקורתית, בדיקה, אתגור והכרעה. כדי שהפיקוח יהיה ממשי נדרש כי למנהל הביקורת תהיה, לכל הפחות, סמכות *override*, גישה למידע רלוונטי, מומחיות מספקת (אוריינות בתחום השימוש ב-AI), תזמון הולם של הפיקוח והבקרה, יכולת תיעוד והימנעות מנטייה לקבל ולאשר באופן לא ביקורתי את תוצרי ה-AI. האדם איננו "פתרון קסם". גם הוא עלול לטעות או לאשרר תוצרי AI ללא ביקורת.

שיקול הדעת המקצועי איננו תוספת חישובית לראיות, אלא חוליית בקרה והכרעה בתוך המהלך כולו על אופן השימוש בהן. תפקידו של מנהל הביקורת איננו להחליף ראיות באינטואיציה, אלא להעריך את הראיות, לבחון את איכותן, לאתגר תוצרים שהופקו בעזרת AI, להחליט מה

מנקודה זו מתחדדת המשמעות הביקורתית של דרישות הראייה בנוגע לתוצרי AI. לפי הנחיות הביקורת וכן לפי תקני ISSAI, ראיות הביקורת צריכות להיות מתאימות ומספיקות: התאמתן נבחנת לפי רלוונטיות, תקפות ומהימנות; ומספיקות עניינה בכמות הראיות הנדרשת לביסוס הממצא. דרישות אלה יוצרות מסנן קפדני במיוחד כאשר נעזרים ב-AI: לא די שתוצר AI "נראה נכון", שאפשר להסבירו או שאפשר לתעד את הפקתו. על מנהל הביקורת לברר אם המידע שהופק בעזרת AI רלוונטי לשאלת הביקורת, אם הוא תקף ביחס לעולם התוכן הנבדק, אם הוא מהימן ומאומת, ואם הוא משתלב בכמות ובאיכות הנדרשות כדי לבסס ממצא. כל עוד אחד מן הרכיבים הללו אינו מתקיים, הפלט נותר במעמדו האנליטי בלבד.

מגבלות אפיסטמיות של תוצרי AI וסיכונים תפעוליים בהפקתם משפיעים אפוא ישירות על כל חוליות שרשרת ההיסק של הביקורת: בשלב הראייה הם פוגעים ביכולת לבסס בסיס אמפירי מהימן; בשלב הממצא הם עלולים ליצור מראית עין של השוואה בין הנורמה לבין מצב בפועל שלא הוכח דיו; בשלב המסקנה הם עלולים להרחיב יתר על המידה מן הממצא אל סיבות, תוצאות או השלכות שאין להן ביסוס מספק בראיות; ובשלב ההמלצה הם עלולים להוביל לפעולה מתקנת שאינה מותאמת לגורם האמיתי לפער. מגבלה אפיסטמית שאינה מטופלת בשלב הראייה אינה נשאת מקומית; היא מקרינה קדימה על כלל השרשרת⁶⁰.

בהקשר זה חשיבותם של כלי תיעוד, אירועי *override*, *cards model*, *datasheets* ורשומות תיקון היא כפולה אך מוגבלת. הם עשויים לסייע בהבנת תנאי ההפקה של תוצר AI, בבדיקת עקיבות ובשחזור תהליך העבודה; הם עשויים גם לסמן נקודות סיכון או מקומות שבהם נדרש אתגור נוסף. אולם הם אינם פותרים את הבעיה הבסיסית של ייחוס מעמד ראייתי לתוצר אנליטי בלא ביסוס מספק. יתר על כן, עצם קיומו של תיעוד עלול לייצר מראית עין של שליטה, גם כאשר בפועל אין בידי מנהל הביקורת יכולת ממשית להעריך את איכות הפלט או את מגבלותיו. בכך מתגלה שוב הפער שבין עקרונות ממשל AI לבין לוגיקת הראיות: שיפור המהימנות, התיעוד והבקרה חשוב, אך אין בכך כדי להבטיח שיפור אוטומטי בתוקף הביקורת⁶¹.

60 משרד מבקר המדינה המדריך לביקורת המדינה, גיבוש מסקנת ביקורת: מהי מסקנת הביקורת? וכן רכיבי מסקנת הביקורת, בעמ' 1 - 2; שם, גיבוש המלצת ביקורת: מהי המלצת ביקורת? וכן דגשים לגיבוש המלצה, בעמ' 1 - 3.

61 Kho, A Framework לעיל ה"ש 38; Mitchell, Model Cards לעיל ה"ש 38; Timnit, Datasheets לעיל ה"ש 38; Mokander, Auditing of AI לעיל ה"ש 20.

62 *human in the loop* (אדם בתוך התהליך) הוא מצב שבו אדם משולב פורמלית בתהליך העבודה עם מערכת AI. במאמר זה המונח מובחן מפיקוח ביקורתי אפקטיבי: עצם נוכחותו של אדם בתהליך אינה מבטיחה בדיקה, אתגור או הכרעה מקצועיים.

לבחינת ראיות הביקורת וכתובתו של דוח הביקורת, אלא הוא גם מי שנדרש לפיקוח אפקטיבי על תהליך העבודה עם AI, גם מי שמעריך וממק את מקומו של מידע שהופק בעזרת AI בתוך שרשרת הביקורת, וגם הנושא באחריות המקצועית לתוצר הסופי.

סינתזה: מסגרת מושגית-נורמטיבית לשימוש תקף ב-AI בתוך המערכת הפורמלית של ביקורת המדינה

מן הדיון המצטבר בפרקים הקודמים מתגבשת כעת הטענה המרכזית של המאמר במלואה: השימוש התקף ב-AI בביקורת המדינה אינו נבחן לפי יעילות הכלי ויכולותיו, ולא לפי איכות התובנה ורמת ההסבריות של התוצר, אלא לפי היכולת לשלב את השימוש במערכת הפורמלית של נורמה - ראייה - ממצא - מסקנה - המלצה. בכך עובר הדיון משיח של עקרונות ממשל AI אל שיח של שימוש תקף בביקורת המדינה - לא הוספת עיקרון חדש לשיח ה-AI, אלא פיתוח "שכבת תרגום" המתווכת בין שפה של עקרונות ונורמות לבין דרישות הראייה, הממצא וההצדקה של ביקורת המדינה, ומסדירה את תנאי המעבר התקפים ביניהן.

שכבת התרגום הזאת היא העיקרון המארגן של המאמר כולו. היא מתחילה בהבחנה בין ממשל AI לבין לוגיקת העבודה וההוכחה של ביקורת המדינה: הראשון מנסח עקרונות, ערכים וחובות; השנייה מגדירה כיצד אפשר לדעת אם אלה אכן התקיימו, באילו קריטריונים ובאילו ראיות. היא נמשכת בהבחנה בין תוצר AI לראייה, בין תובנה לממצא, ובין אחריותיות והסבריות לתיקוף וביסוס ראייתי. היא מחייבת שלא להותיר מושגים שונים מעולם ממשל AI ברמת עקרונות כלליים בלבד, אלא להתאימם לעולם הביקורת: לתרגמם, לפי העניין, לקריטריוני ביקורת אופרטיביים, לתנאי בדיקה, לדרישות תיעוד ועקיבות ולשיקולים של עיגון ראייתי והצדקה. בכך היא ממירה את הפער שבין השיח הנורמטיבי הרחב על AI לבין הלוגיקה הפורמלית של ביקורת המדינה למבנה מעבר שיטתי.

במונחים מוסדיים, המסגרת המוצעת קובעת כי מקומו של AI בביקורת הוא תמיד נגזר-תהליך, משולב בתהליך, ולא מחליף תהליך. AI אינו חוליה בשרשרת הביקורת, אלא כלי ששימוש בו ממוקם ביחס לשרשרת נורמה - ראייה - ממצא - מסקנה - המלצה, ובכפוף לה. AI יכול לסייע בתכנון, בארגון חומר, בזיהוי דפוסים, בסיווג מסמכים, באיתור כיווני בדיקה ובסיוע טקסטואלי זהיר; אך ככל

מאומת ומה נדחה, ולהצדיק כיצד מידע שהופק בעזרת AI נבחן, אומת והוצלב, ובאיזה אופן יש לשלב אותו בראיות הביקורת שעל בסיסו יגובש ממצא. במובן זה, פיקוח אנושי הוא רכיב אינהרנטי של שליטה אפיסטמית: הוא נדרש לא מפני שהטכנולוגיה "מסוכנת" בלבד, אלא מפני שללא שיקול דעת מקצועי מתועד ומנומק אין כלל דרך תקפה לעבור מתוצר AI אל הראייה, ומן הראייה אל הממצא.

המשמעות היא שיש להבחין בין פיקוח אנושי לבין אחריות פורמלית גרידא. העובדה שאדם חתם, אישר או היה "בתהליך" איננה מוכיחה שהופעל שיקול דעת ביקורתי. בדומה לכך, אחריותו של מנהל הביקורת לדוח הסופי אינה מתקיימת מעצם היותו בעל הסמכות, אלא מכוח היכולת להצדיק את ההליך שהוביל אל הממצא, אל המסקנה ואל ההמלצה. לכן **פיקוח אנושי על AI בביקורת המדינה הוא מנגנון של אתגור ובחינה ביקורתית: היכולת לשאול מה מקור הפלט, מה נבדק, מה לא נבדק, מהי דרגת הביטחון, אילו הסברים ותיעודים זמינים, ומה עדיין חסר כדי לבחון אם מידע שהופק בעזרת AI יכול להשתלב בראיות הביקורת.** רק במובן זה נעשה הפיקוח האנושי חלק אינהרנטי מן ההליך, ולא רק נספח של עקרונות ממשל AI.

מעמד זה של שיקול הדעת המקצועי מתחדד בייחוד בשלב המעבר מממצא למסקנה וממסקנה להמלצה. מי שמכריע מהי משמעותו של הממצא, מהו הגורם לפער ואיזו המלצה נגזרת ממנו אינו הכלי, אלא מנהל הביקורת. כלי AI יכול אולי להציע דרכי ארגון, לזהות תבניות, להעלות אפשרויות ניסוח או להצביע על כיווני בדיקה; אך ההכרעה אם הוכחה הסיבה, אם בוססו התוצאות ואם המלצה מסוימת מוצדקת היא הכרעה ביקורתית מקצועית שעל מנהל הביקורת לעמוד מאחוריה.

יתר על כן, הנחיות מבקר המדינה מבהירות כי הביקורת אינה מחליפה את שיקול הדעת של הגוף המבוקר, וכי המלצות הנוגעות לעניינים מקצועיים או לעניינים של סבירות מחייבות פגם מוכח בשיקול הדעת או דרך מדידה המאפשרת להוכיח חוסר סבירות או תוצאות שליליות בלתי מידתיות⁶³. קביעה זו רלוונטית גם לשימוש ב-AI על ידי הביקורת עצמה. אם הביקורת אינה רשאית לעקוף את שיקול דעתו של הגוף המבוקר בלי ביסוס ראייתי מספק, קל וחומר שאין היא רשאית לעקוף את שיקול דעתה שלה באמצעות מידע שהופק בעזרת AI. שיקול הדעת המקצועי של מנהל הביקורת הוא אפוא לא רק מנגנון פנימי של שליטה, אלא גם תנאי מוסדי לגיטימיות של דוח הביקורת.

סיכום של דבר, מעתה מנהל הביקורת אינו רק האחראי על יישום וביצוע של שלבי עבודת הביקורת, אינו רק האחראי

63 משרד מבקר המדינה המדריך לביקורת המדינה, עבודת השדה: גיבוש ממצא ביקורת, בעמ' 3; שם, גיבוש המלצת ביקורת, בעמ' 2 - 3.

סיכום והשלכות למחקר המשך

מאמר זה ביקש להראות כי שאלת השימוש ב-AI בביקורת המדינה איננה שאלה של יעילות טכנולוגית או עמידה כללית בעקרונות. השאלה המכריעה היא אם וכיצד אפשר לשלב שימוש כזה במערכת הפורמלית של נורמה - ראיה - ממצא - מסקנה - המלצה, בלי לפגוע בלוגיקת ההוכחה ובתוקף הראייתי שעליהם נשען דוח הביקורת. טענתנו היא כי מסגרות ה-AI הכלליות חשובות, אך אין בהן דין; נדרשת שכבת תרגום המאפשרת מעבר תקף בין שיח ה-AI לבין לוגיקת הביקורת. בהיעדרה, גם תוצר אנליטי מועיל, מוסבר או מתועד היטב אינו נעשה ראיה, אינו נעשה ממצא, ואינו יכול לשמש בסיס ישיר למסקנה או להמלצה.

במובן זה, מבחינה מושגית אפשר להבחין בין שיח ה-AI לבין לוגיקת הביקורת, ולטעון כי ממשל AI אינו לוגיקת העבודה וההוכחה של ביקורת המדינה. מבחינה אפיסטמית אפשר להגדיר את תנאי המעבר התקפים מתוצר אנליטי אל מערך ראיות שעל בסיסו בלבד ראוי לגבש ממצא, מסקנה והמלצה. מבחינה מוסדית-ביקורתית אפשר למקם את מנהל הביקורת לא כ"חותמת אנושית" על תוצרי AI, אלא כאחראי על פיקוח ביקורתי והכרעה מקצועית להערכת הראיות והתוצרים המופקים, לאתגורם ולהצדקת מקומם בדוח הביקורת, ולהבהיר כי פיקוח אנושי איננו נוכחות טכנית בלבד, אלא שיקול דעת מקצועי הנושא באחריות לתוצר הסופי. על יסוד זה הוצגה המסגרת המוצעת לא כרשימת עקרונות, לא כמודל טכנולוגי ולא כמודל הפעלה (Operating Model), אלא כמסגרת מושגית-נורמטיבית הקובעת את תנאי המעבר התקפים ומתווכת בין עקרונות כלליים של שיח ה-AI לבין עבודת הביקורת עצמה: עיגון בנורמות, בחינה על בסיס ראיות, גיבוש ממצאים, ניסוח מסקנות והצגת המלצות.

מתוך כך מאמר זה פותח שני כיוונים משלימים של מחקר עתידי. הראשון הוא פיתוח מודל הפעלה לשימוש ב-AI לאורך מחזור העבודה של ביקורת המדינה, שיתרגם את המסגרת המושגית-נורמטיבית למבנה עבודה סדור לאורך שלבי התכנון, עבודת השדה, גיבוש הדוח, הפרסום והמעקב. השני הוא פיתוח מסגרת פיקוח ובקרה (Control Framework) שתגדיר מנגנוני אימות, תיעוד, עקיבות ובקרה לעבודת הביקורת הנתמכת ב-AI. שני כיוונים אלה אינם מחליפים את המסגרת שהוצעה כאן, אלא נבנים ונשענים עליה: בשים לב לתנאי התוקף לשימוש ב-AI בביקורת המדינה שהוגדרו במאמר זה, המחקר העתידי נדרש לברר כיצד ליישםם יישום שיטתי, מבוקר ועקיב בתוך פרקטיקת הביקורת עצמה.

שההליך מתקדם לעבר גיבוש ממצא, מסקנה והמלצה, כך מתחדדים ומתקשחים תנאי התרגום, האימות וההכרעה. אין קפיצה ישירה מן הפלט אל הממצא; אין שקילות בין תיעוד לבין ראיה; אין די בהסבר של התוצר כדי לבסס ליקוי; ואין בפיקוח אנושי טכני כדי להכשיר שימוש שאינו עומד בדרישות ההוכחה. לפיכך המסגרת המושגית-נורמטיבית דנא איננה מבנה "מעודד חדשנות", אלא מבנה תוחם: היא מגדירה מתי, כיצד ובאילו תנאים שימוש ב-AI יכול להיות קביל בעבודת הביקורת.

על רקע זה מתברר כי תרומתה של המסגרת המוצעת במאמר היא משולשת ומשלימה. ראשית, זוהי תרומה מושגית, משום שהיא מחדדת את ההבחנה בין שיח ממשל AI לבין לוגיקת העבודה וההוכחה של ביקורת המדינה. שנית, זוהי תרומה אפיסטמית, משום שהיא מגדירה את תנאי המעבר התקפים מתוצר אנליטי, ובכלל זה תוצר שנעזר ב-AI, אל מערך של ראיות ביקורת שעל בסיסו בלבד אפשר לגבש ממצא, מסקנה והמלצה. שלישית, זוהי תרומה מוסדית-ביקורתית, משום שהיא ממקמת את מנהל הביקורת לא כחותמת אנושית על תוצרי המערכת, אלא כמנגנון ההכרעה המקצועי האחראי להערכת הראיות, לאתגורם ולהצדקת מקומן בדוח הביקורת.

לאור האמור במאמר זה, אפשר להציע שלוש תזוזות משולבות. ראשית, תזוזה מושגית: משיח כללי של ממשל AI, אחריותיות ומהימנות AI אל הבהרת מיקומם ביחס ללוגיקת הביקורת. שנית, תזוזה אפיסטמית: מדיון בשאלה מה ראוי או יעיל אל דיון בשאלה מה אפשר לבסס, לאמת ולהצדיק במסגרת דוח ביקורת מבוסס ראיות. שלישית, תזוזה מוסדית-ביקורתית: מתפיסה של AI כתוספת טכנולוגית חיצונית אל מיקומו הכפוף לשרשרת הפורמלית של נורמה - ראיה - ממצא - מסקנה - המלצה. תזוזות אלה אינן עומדות זו לצד זו; הן כרוכות זו בזו. רק כאשר הדיון המושגי מתורגם לדיון אפיסטמי, ורק כאשר הדיון האפיסטמי ממקם בתוך המבנה המוסדי של הביקורת, אפשר לדבר על שימוש תקף ב-AI בביקורת המדינה.

לכן יש לתרגם ולהתאים שיח נורמטיבי וטכנולוגי למבנה מושגי-נורמטיבי הניתן לבחינה, לעיגון ראייתי ולהצדקה במסגרת דוח ביקורת המדינה. זוהי שכבת התרגום שהייתה חסרה: לא עוד שיח כללי על AI אחראי, ולא עוד הסתפקות בתייעוד, בהסבר או בפיקוח אנושי פורמלי, אלא ארכיטקטורה של מעבר שבה נבחן באילו תנאים מידע שהופק בעזרת AI יכול להשתלב בראיות הביקורת, כיצד אפשר לבסס ממצא, מה נדרש כדי לפרש אותו כמסקנה, ואיך אפשר לגזור ממנו המלצה צופה פני עתיד בלי לפגוע בלוגיקת ההליך. במובן זה, המסגרת המוצעת במאמר איננה רק אפשרות פרשנית אחת בין אחרות; היא תנאי להמרת הדיון ב-AI מרמת עקרונות לרמת תיקוף, ומכאן גם לגיטימיות של השימוש בו בביקורת המדינה.



סוכני בינה מלאכותית בשירות ביקורת המדינה

חגי לדרר ואופיר מזרחי

מר חגי לדרר הוא סגן מנהל אגף בתחום הנתונים והסייבר באגף התמיכה המקצועית שבמשרד מבקר המדינה ונציב תלונות הציבור.

מר אופיר מזרחי הוא אחראי לפיתוח למידה באגף הלמידה והפיתוח הארגוני שבמשרד מבקר המדינה ונציב תלונות הציבור.

המאמר בקצרה

- שילוב סוכני AI מבוססי ידע ארגוני עשוי לייעל את עבודת הביקורת, להבטיח אחידות ולשפר את איכות הדוחות.
- פיתוח הסוכנים מבוסס על מתודולוגיית RAG, המחייבת בקרת נתונים קפדנית של מומחי תוכן כדי למנוע טעויות של הסוכנים.
- העובדים צופים שה-AI יעמיק את עבודת הביקורת ללא פגיעה בכוח האדם, אך ידרוש רכישת מיומנויות חדשות.
- אף שתוצרי הסוכנים נתפסים כאיכותיים ומדויקים, מתגלה קושי ליישם אותם באופן פרקטי בעבודה השוטפת¹.
- הצלחת הכלים תלויה בגישור על הפער היישומי, בצמצום סיכונים כמו אובדן מיומנויות אנושיות ובהשקעה בלמידה ארגונית רציפה.

מבוא | רקע תיאורטי | מתודולוגיית RAG לבניית סוכני בינה מלאכותית לצורכי ניהול ארגוני |
הערכת תוצרים של סוכני בינה מלאכותית שפותחו במשרד מבקר המדינה | מהלך המחקר |
תוצאות | סיכום

1 פער יישומי.

מבוא

העבודה בעשור הקרוב, לקראת שנת 2030 צפויה טרנספורמציה משמעותית שתונע משילוב עמוק של בינה מלאכותית בסביבת העבודה ובהכשרה המקצועית. הבינה המלאכותית צפויה לתרום להשגת עבודה יעילה וצמיחה כלכלית באמצעות שיפור ניכר של הגברת מוכנות כוח האדם לתפקידים שונים. במקביל, אוטומציה ומערכות רובוטיקה יבצעו משימות רוטיניות וחוזרות על עצמן, מה שיסחרר את העובדים להתמקד בפעילויות מורכבות יותר הדורשות חשיבה מתקדמת.⁴

בשל ההתפתחות המהירה של הטכנולוגיה, למידה לאורך החיים (Lifelong learning) ופיתוח רציף של מיומנויות יהפכו לתנאים הכרחיים להצלחה בשוק העבודה. פלטפורמות מבוססות בינה מלאכותית יזהו פערי ידע אצל עובדים ויספקו להם הכשרה מותאמת אישית, ובכך יקדמו רכישת מיומנויות חדשות וחיזוק מיומנויות קיימות (Upskilling, Reskilling) בהתאם לדרישות המשתנות של התעשייה. נוסף על כך, מוסדות ההכשרה יוכלו לעדכן את תוכניות הלימוד שלהן בזמן אמת באמצעות שימוש בניתוח נתונים של הבינה המלאכותית, מה שיבטיח צמצום של הפערים בין הנלמד באקדמיה לבין הצרכים המעשיים בשטח.⁵

סוכני בינה מלאכותית בארגונים

סוכני בינה מלאכותית (AI agent) הם טכנולוגיות המבצעות משימות קוגניטיביות שבעבר דרשו ידע, תפיסה ויכולות אנושיות. סוכנים אלה פועלים ברמות משתנות של אוטונומיה ומסוגלים לקיים אינטראקציה ולעבוד בשיתוף פעולה עם בני אדם בתוך צוותים. יש להם יכולת להתפתח, ללמוד מתוך נתונים ולצבור ניסיון בביצוע משימות.⁶ הם מתוכננים לפעול גם באופן עצמאי ולקבל החלטות מורכבות על בסיס נתונים המתקבלים בזמן אמת ללא התערבות אנושית. הם מסוגלים אף לשתף פעולה במערכות מרובות-סוכנים (Multi-agent systems) ולהשתמש באלגוריתמים

עולם העבודה צפוי לעבור בעשור הקרוב טרנספורמציה עמוקה המונעת משילוב נרחב של טכנולוגיות בינה מלאכותית (להלן גם - AI) בסביבות העבודה ובהכשרה המקצועית.² שילוב טכנולוגי זה יוביל לאוטומציה של משימות רוטיניות ושגרתיות ויאפשר לעובדים להשתחרר מביצוע מטלות טכניות כדי להתמקד בפעילויות מורכבות יותר הדורשות חשיבה אנליטית וביקורתית.³

על רקע ההתפתחויות העולמיות המהירות של כלי הבינה המלאכותית והשימוש בהם, מאמר זה מתמקד במשרד מבקר המדינה ונציב תלונות הציבור שבו פותחו כחלק מתנופת חדשנות אסטרטגית מספר סוכני בינה מלאכותית ייעודיים אשר יכולים לשמש כעוזרים מקצועיים עבור עובדי המשרד. הסוכנים שנבחנו הם **סוכן המדריך לביקורת**, השואב את ידיעותיו מתורת העבודה הסדורה במדריך הביקורת ודוחות העבר ו**סוכן עריכת הלשון**, המבוסס על ההנחיות המשרדיות בנושא ובהן אוגדן הכתיבה. שני הסוכנים נועדו לתמוך באופן ישיר בליבת העשייה המשרדית. מטרתו המרכזית של המחקר שבוצע במסגרת מאמר זה היא להעריך באופן ביקורתי אם העוסקים בביקורת המדינה במשרד סבורים כי השימוש בכלי בינה מלאכותית בכלל, ובפרט בסוכני בינה מלאכותית ייעודיים אלו - המוזנים בידע הארגוני והמתודולוגי הייחודי של המשרד, מסוגלים הלכה למעשה לייעל את עבודת הביקורת, להבטיח את האחידות המתודולוגית הרצויה ולשפר את איכות הניסוח של דוחות הביקורת המפורסמים לציבור.

רקע תיאורטי

עתיד עולם העבודה עד שנת 2030

על פי יעדי הפיתוח בר-הקיימא של האו"ם לעולם

Febri Prasetya, Aprilla Fortuna, Agariadne Dwinggo Samala, Dinda Khaira Latifa, Welli Andriani, Utari Akhir Gusti, Muhammad Raihan, Santiago Criollo-C, Deniz Kaya, Juan Luis Cabanillas García, Harnessing Artificial Intelligence to Revolutionize Vocational Education: Emerging Trends, Challenges, and Contributions to SDGs 2030, 11 Soc. Sci. & Hum. Open 101401 (2025), <https://doi.org/10.1016/j.ssaho.2025.101401>

Mohammad Hosseini, Maya Murad & David B. Resnik, Benefits and Risks of Using AI Agents in Research, 56(1) Hastings Center Report 13 (2026), <https://doi.org/10.1002/hast.70025>

4 Prasetya et al., Harnessing Artificial Intelligence, לעיל ה"ש 2.

5 שם.

6 Patrick Figge, Edward Anderson & Kyle Lewis, AI-Human Learning Systems: Investigating the Strategic Role of AI for Organizational Learning 24(2) Strategic Organization (2025) <https://doi.org/10.1177/14761270251385860>



כריית נתונים¹³ שנועדה לחשוף מידע מורכב, משפרות את היעילות והאפקטיביות של בדיקות מבססות ומפחיתות טעויות אנוש. למרות הפוטנציאל הגדול לשיפור של איכות הדיווח הכספי ושל יעילות הביקורת, מנהלים פיננסיים בחברות המבוקרות מביעים לעיתים דעות מעורבות לגבי ההשפעה הישירה של טכנולוגיות אלה על החברות שלהם. מצד אחד, מנהלים מאמינים שהבינה המלאכותית תשפר את איכות הביקורת, תגביר את שביעות הרצון ותפחית טעויות אנוש. מצד שני, יש ספקנות רבה בעניין השאלה אם התייעלות זו תוזיל את עלות הביקורת או תצמצם את זמן העבודה והמאמץ הנדרשים מצוותי הכספים של החברה המבוקרת. יתרה מכך, מחקרים מראים כי היכולת של ביקורת מבוססת AI לשפר בפועל את איכות הדיווח תלויה בנכונות של הנהלת החברה לאמץ את הממצאים. נכונות זו מתגברת באופן ניכר כאשר החברה המבוקרת עצמה משתמשת בבינה מלאכותית בתהליכי הפנימיים ורואה בה מקור מידע מדויק ואמין.¹⁴

סיכונים והזדמנויות בשימוש בבינה מלאכותית

בשימוש בכלי AI טמונות הזדמנות, אך גם סיכונים שיכולים להיות בעלי משמעות מעשית. להלן בלוח 1 ריכוז ההזדמנויות והסיכונים בשימוש בכלי AI בחקירת נושאים מורכבים.

מתקדמים של למידת מכונה.⁷ דוגמאות מעשיות לכך הן סוכנים המסוגלים לתמלל פגישות בזמן אמת ולהפיץ את המידע לאנשים רלוונטיים⁸ וכן סוכנים המנטרים עומס קוגניטיבי ומנתחים מדדים שונים כדי להתאים אישית את סביבת העבודה ולהפחית הסחות דעת אצל העובדים.⁹

אחת הבעיות המורכבות בארגונים היא "התמחות יתר" (Overspecialization) - מצב שבו עובדים מפתחים מומחיות ספציפית עד שיכולתם לתקשר ולשלב ידע עם אחרים נפגעת. סוכני AI המבוססים על הידע הפנים ארגוני יכולים לתת מענה לכך. ה-AI מנהל ומשמר את הידע המשותף (הקולקטיבי) של הארגון ומספק לעובדים את ה"מרחב הקוגניטיבי" הנדרש כדי להעמיק את מומחיותם הייחודית וללמוד תחומים חדשים במהירות לצד שמירה על תקשורת זורמת וביצועים גבוהים.¹⁰ לצד זה, היכרות הסוכנים עם סביבת העבודה מאפשר להם לספק לעובדים התערבויות מותאמות אישית, כגון מענה לבעיות על בסיס המידע הארגוני.¹¹

שימושים בבינה מלאכותית בביקורת

השימוש בבינה מלאכותית בתחומי ביקורת כגון ביקורת פנים וביקורת חשבונאית הולך ומתרחב. משרדי רואי חשבון, למשל, נעזרים בטכנולוגיות מתקדמות כמו למידת מכונה ועיבוד שפה טבעית כדי לבצע מטלות שבאופן מסורתי דרשו אינטליגנציה אנושית. אחד השימושים המרכזיים של בינה מלאכותית בביקורת הוא ביצוע פונקציות מורכבות כמו הערכת סיכונים, קביעת נהלים אנליטיים וביצוע הערכות שווי. טכנולוגיות אלה מאפשרות לרואי החשבון המבקרים לבחון אומדנים חשבונאיים מורכבים וסובייקטיביים בצורה מדויקת ואמינה יותר, בין היתר משום שמערכות ה-AI מסוגלות לשלב ולעבד כמות נתונים גדולה בהרבה מזו שמומחים אנושיים מסוגלים לנתח.¹²

טכנולוגיות AI מסייעות לבצע חישוב של הכנסות, מאפשרות

7 Ravirajan Karunanithi & Arvind Sundarajan, A. Enhancing Workplace Productivity and Well-being Using AI Agent ARXIV (January 2025) [arXiv:2501.02368](https://arxiv.org/abs/2501.02368) [cs.AI].

8 Figge, Anderson & Lewis, AI-Human Learning Systems, לעיל ה"ש 6.

9 Ravirajan & Sundarajan, Enhancing Workplace, לעיל ה"ש 7.

10 Figge, Anderson & Lewis, AI-Human Learning Systems, לעיל ה"ש 6.

11 Ravirajan & Sundarajan, Enhancing Workplace, לעיל ה"ש 7.

12 Cassandra Estep, Emily Elaine Griffith & Nicole L. MacKenzie, How Do Financial Executives Respond to the Use of Artificial Intelligence in Financial Reporting and Auditing? 29(3) Rev. of Accounting Stud. 2798 (2024), <https://doi-org.elib.openu.ac.il/10.1007/s1142-023-09771-y>

13 תהליך של ניתוח מאגרי מידע ענקיים במטרה לגלות בהם דפוסים, מגמות וקשרים שיסייעו בהסקת מסקנות ובקבלת החלטות.

14 Estep, Griffith & MacKenzie, How Do Financial Executives Respond, לעיל ה"ש 12.

התחום	הסיכונים	ההזדמנויות
יעילות העבודה וביצוע המשימות	הזיות ¹⁵ וטעויות עובדתיות המערכות מועדות לטעויות, היות, חוסר דיוק עובדתי ושגיאות. לעיתים המאמץ הנדרש לאיתור ותיקון של טעויות אלה עולה על החיסכון הראשוני בזמן.	ביצוע משימות רוטניות ללא הפסקה כלי AI יכולים לבצע כתיבה, עריכה, קידוד, ניתוח נתונים ותכתובות בדואר אלקטרוני ולאפשר לעובדים להתפנות למשימות הדורשות תובנות עמוקות יותר. נוסף על כך, לכלים אלה יש מהירות עיבוד אדירה (יכולת לנתח כמויות עצומות של טקסטים ומסמכים).
אמינות ואחריות	פער אחריות (Responsibility Gap) הסתמכות יתר על בינה מלאכותית עלולה לגרום לעובדים להימנע מלבקר את עבודת הבינה המלאכותית. במקרה של כשל או החלטה שגויה קשה לקבוע מי נושא באחריות המשפטית והמוסרית, מכיוון שלמכונה אין סוכנות מוסרית (Moral agency) והאדם לא תמיד מסוגל לפקח עליה כראוי.	הפחתת טעויות אנוש צמצום שגיאות הנובעות מעייפות, מחוסר תשומת לב, מביצוע לקוי או מאי-ציות לנוהלי בטיחות.
מיומנויות ויכולות אנליטיות	"אובדן מיומנויות (Deskilling)" העברת מטלות קוגניטיביות למכונה עלולה לפגוע ביכולות חיוניות של עובדים, כמו חשיבה ביקורתית, תקשורת ודמיון. בעקבות זאת עובדים עלולים לאבד את היכולת המעשית לאמת ולבדוק את תוצרי הבינה המלאכותית כנדרש.	העמקת המיקוד המחקרי היכולת להריץ תהליכים מרובי-שלבים וניתוחים מורכבים כך שיתנהלו באופן עצמאי מאפשרת לעובדים להתמקד בחשיבה על פתרון הבעיה ולא בביצוע הטכני.
תעסוקה והכשרת עובדים	אובדן משרות התחלתיות אוטומציה של משימות מחקר שגרתיות צפויה לבטל משרות של סטודנטים ומתמחים. תפקידים אלה חיוניים להתפתחות המקצועית של העובדים ולהפיכתם למומחים עצמאיים בהמשך.	פתיחת אפיקים חדשים כלי AI עשויים לייצר מקומות עבודה חדשים באמצעות פתיחת תחומי מחקר ויצירת תפקידים חדשים שטרם היו קיימים.
אתיקה ויושרה	פעולות לא אתיות מצד ה-AI המכונה עלולה לבדות נתונים (הזיות).	עקיפת מגבלות אתיות בניסויים כלי AI מסוגלים לדמות התנהגות ותגובות של נבדקים אנושיים (ליצור "תאומים דיגיטליים" ברמת דיוק של 85%), מה שמאפשר לבצע מחקרים שביצועם בהשתתפות בני אדם יהיה יקר, לא מעשי או לא אתי.

15 נטייה של מודלי בינה מלאכותית לספק תשובות שגויות בביטחון רב.

האצת גילויים ותבניות כלי AI יכולים לייצר השערות מחקר חדשות, למצוא קשרים נסתרים בספרות המקצועית ולתכנן מודלים של מערכות מורכבות.	ידע שאינו ניתן לאימות: (Unverifiable) ככל שכלי AI יהפכו למתקדמים יותר הם יהיו עשויים לייצר ממצאים, תיאוריות או מודלים שיהיו כה מורכבים עד שהמוח האנושי לא יהיה מסוגל להבין או לאמת אותם באופן מוחלט. מה שמעלה שאלות יסוד לגבי טיב הידע וההסתמכות עליו.	אופי הידע והממצאים
	AI ככלי ריגול ומעקב שימוש של עובדים בכלי AI שלא בהתאם להנחיות אבטחת המידע, בגלוי או בסתר (Shadow AI), עלול לחשוף את הארגון לסיכונים כתוצאה מיכולותיה המתקדמות של הבינה המלאכותית, כתוצאה מיכולותיה המתקדמות של הבינה המלאכותית והחשיפה של נתונים אישיים, רגישים או סודיים.	פרטיות ואבטחת מידע

על פי נתוני *Benefits and Risks of Using AI Agents in Research* ו-*The shape of the shadow of The Thing*, בעיבוד משרד מבקר המדינה.

מתודולוגיית RAG לבניית סוכני בינה מלאכותית לצורכי ניהול ידע ארגוני

המתודולוגיה המרכזית לבניית סוכני בינה מלאכותית (ומודלי שפה בפרט) המבוססים על ידע ארגוני נקראת RAG (Retrieval-Augmented Generation). מתודולוגיה זו נועדה לפתור את המגבלות של מודלי שפה גנריים (LLMs) - שלרוב חסרים ידע ספציפי לארגון, מוגבלים בכמות המידע שהם מסוגלים לעבד בבת אחת ונוטים לספק תשובות שגויות בביטחון רב ("הזיות") כאשר הם נשאלים על מידע שאינו מוכר להם. מתודולוגיית RAG משלבת בין מודל השפה לבין מאגר ידע ארגוני חיצוני ומאפשרת למודל לשלוף מידע ארגוני רלוונטי ולבסס עליו את תשובותיו. להלן בתרשים 1 מפורטים ארבעת השלבים העוקבים של מתודולוגיית ה-RAG לבניית סוכני בינה מלאכותית¹⁶.

1. חילוץ נתונים (Data Extraction) ממערכות המידע הארגוניות: בשלב זה המערכת אוספת את המידע הארגוני הרלוונטי ממגוון רחב של מקורות נתונים, הן מקורות מובנים כגון בסיסי נתונים וטבלאות והן

תרשים א': השלבים העוקבים של מתודולוגיית RAG



Leopold Müller et al., Data Quality Challenges in Retrieval-Augmented Generation, ARXIV (October 2025) [arXiv:2510.00552](https://arxiv.org/abs/2510.00552) [cs.AI]. 16

מקורות שאינם מובנים כגון קובצי מלל. האתגר המרכזי בשלב זה הוא שמירה על דיוק הנתונים ואיכותם, לדוגמה טיפול בכפילויות, זיהוי נתונים סותרים או לא רלוונטיים או עמידה בדרישות פרטיות¹⁷.

הבסיס לסוכן אמין הוא איכות הידע שמוזן אליו. שלב זה דורש שיתוף פעולה הדוק בין גורמים טכנולוגיים לגורמים מקצועיים. הזנת המידע ואפיון המקורות אינם נתונים לאחריותה של מחלקת מערכות מידע בלבד, אלא גם לאחריותם של מומחי התוכן בארגון. לדוגמה, אם מפתחים סוכן שנועד לסייע בתהליכי ביקורת, הגורמים שצריכים לאשר את המידע הם גם גורמי הבקרה ופיתוח המתודולוגיות בחטיבות המטה וגם מנהלים בכירים מתחום הביקורת. תפקידם לקבוע אילו מסמכים הם "אמת הארגונית" (למשל המדריך המקצועי לביקורת או אוגדן ההנחיות לעריכת לשון) ולסנן טיטות ישנות או נהלים שבוטלו.

2. התמרת נתונים (Data Transformation): חרף ההתקדמות בהתפתחות כלי הבינה המלאכותית, עדיין קיימות הגבלות במודלים, כגון יכולת מוגבלת לעבד טקסטים ארוכים. כדי להתמודד עם מגבלה זו יש צורך לבצע תהליך הכולל חלוקה של הנתונים והמסמכים למקטעים קטנים ונוחים יותר לעיבוד (Chunking). לאחר מכן המקטעים הללו מומרים לייצוגים מתמטיים-וקטוריים (Embeddings) ומאוחסנים בתוך "מסד נתונים וקטורי" ייעודי המאפשר שליפה יעילה. מתודולוגיה נכונה דורשת למצוא את אסטרטגיית החלוקה המדויקת. חלוקה שגויה עלולה לקטוע משפטים באמצע, לאבד את הקשר המידע או להתעלם ממידע חשוב שמסתתר באמצע המקטע¹⁸.

משכך, המידע אינו יכול להתווסף כפי שהוא מופיע במערכות המידע המשרדיות. על מומחי התוכן לוודא שהמסמכים מובנים בצורה שמודל השפה יוכל להבין ולפרש בצורה מהימנה ובלי לפגוע בהקשר של המידע. יש לחלק את המידע למקטעים (Chunking), לדוגמה לפי סעיפים ותתי-סעיפים, בצורה היררכית.

3. הנחיה וחיפוש (Prompt & Search): כאשר משתמש בארגון מזין שאילתה למערכת, השאילתה עצמה מומרת לייצוג וקטורי¹⁹. לאחר מכן, המערכת סורקת את מסד הנתונים הווקטורי ומשתמשת באלגוריתמים של חישוב דמיון (Similarity matching) כדי לאתר ולשלוף את המקטעים הרלוונטיים ביותר לשאלה. בשלב זה, חשוב להבטיח את רמת הדיוק וכן לוודא כי תתאפשר שליפת נתונים אך ורק מתוך מסמכים ומקורות מידע שהעובד מורשה להיחשף אליהם. יש להימנע מיצירת מנגנון שיחשוף מידע ארגוני סודי בכני עובדים שאינם מורשים לצפות בו²⁰.

4. יצירת התשובה (Generation): בשלב הסופי והאחרון המקטעים שנשלחו ממאגר הידע מאוגדים יחד עם שאלת המשתמש המקורית, ומתווספת אליהם הנחיית רקע (System Prompt) מוגדרת מראש שמנחה את המודל כיצד לענות, באיזה סגנון ותחת אילו מגבלות. כל המידע המשולב מוזן למודל השפה שמעבד ומשלב אותו כדי לייצר עבור המשתמש תשובה קוהרנטית, מדויקת ומותאמת הקשר²¹.

כדי לכתוב את הנחיות הרקע באופן מיטבי, על מפתחים של סוכני AI לכלול את המידע בצורה הבאה:

- **הגדרת תפקיד וטון**²²: יש להגדיר את האישיות של הסוכן. למשל: "אתה עוזר מקצועי המשמש מנטור לעובדי הארגון. עליך לספק תשובות ענייניות, מדויקות ומבוססות על עובדות בלבד". תפקיד הנחיית המערכת הוא לכוון באופן רחבי את ההתנהגות וסגנון התגובה (הטון) של מודל השפה, בעוד שפרומפט²³ גיל מייצג את השאילתה או השאלה הספציפית שהמשתמש מזין למערכת באותו הרגע.
- **עקרון ההיצמדות לידע**: ההנחיה החשובה ביותר במערכת RAG היא לאסור על המודל להשלים פערי מידע ממקורות שלא הוזנו לו מראש. ההנחיה צריכה להיות מפורשת: "בסס את תשובתך אך ורק על המקטעים המצורפים. אם המידע הדרוש

17 שם.

18 שם.

19 ייצוג מספרי המשקף את המשמעות וההקשר של הטקסט כדי שהמערכות הממוחשבות יוכלו לעבדו באופן מיטבי.

20 Müller et al., Data Quality Challenges, לעיל ה"ש 16.

21 שם.

22 סגנון התגובה ודפוס ההתנהגות של מודל השפה בעת מתן המענה למשתמש.

23 שאילתה או ההנחיה שהמשתמש מזין למודל השפה.

שהסוכן סיפק תשובה לא מדויקת או שלף נוהל מיושן, המשוב חוזר למנהלי הידע לצורך עדכון מאגר הנתונים, מה שמבטיח שיפור מתמיד של המודל.

הערכת תוצרים של סוכני בינה מלאכותית שפותחו במשרד מבקר המדינה: ניתוח אמפירי של עמדות עובדים במשרד

כחלק מתנופת החדשנות וייעול תהליכי העבודה, פותחו במשרד מבקר המדינה ונציב תלונות הציבור מספר סוכני בינה מלאכותית (AI Agents) ייעודיים אשר טרם הופצו לכלל העובדים במשרד. סוכנים אלו נועדו לשמש עוזרים מקצועיים לעובדים. הם מבוססים על מתודולוגיות העבודה והמדריכים המקצועיים של המשרד ונועדו לתמוך באופן ישיר בליבת העשייה המשרדית - החל בביצוע עבודת הביקורת, דרך ייעול הליכי בירור התלונות ועד לבקרה קפדנית על הדוחות. הפעלתם תאפשר לצוותים לעבד ולנתח כמויות מידע גדולות במהירות, לשלוף הנחיות רלוונטיות בזמן אמת מתוך הידע הארגוני ולהבטיח רמת דיוק, עקביות ואיכות גבוהה לאורך כל שלבי הביקורת ובירור התלונות. להלן תיאור שניים מן הסוכנים שפותחו:

1. **סוכן המבוסס על המידע המופיע במדריך המקצועי לביקורת ועל דוחות ביקורת שפורסמו בעבר** - הסוכן נועד לסייע לעובדי משרד מבקר המדינה בביצוע עבודת הביקורת הלכה למעשה. המטרה המרכזית היא להבטיח שעבודת הביקורת תיעשה באופן המדויק, המקצועי והאיכותי ביותר ועל פי הנחיות המדריך לביקורת. כדי לסייע לעובדי המשרד, הסוכן ניזון מתשתית מקצועית ומעשית רחבה המורכבת משני נדבכים עיקריים: הנדבך הראשון הוא המדריך לביקורת, אשר "מטרותיו העיקריות הן לסייע בהבטחת איכות הביקורת בכל שלביה וליצור תורה סדורה בעניין ביקורת המדינה"²⁶. המדריך לביקורת הוא "ספר ההפעלה" המחייב שממנו שואב הסוכן את הכללים לכל שלבי הביקורת - החל מהשיטות לאיסוף מידע, דרך אופן הביסוס והניסוח של הממצאים ועד להתנהלות הרגישה מול המבוקרים בשלב קבלת התגובות. הנדבך השני שממנו הסוכן שואב את הידע הוא מאגר

למענה על השאלה אינו מופיע במקטעים אלו, עליך לענות באופן חד-משמעי: 'המידע אינו קיים במאגר הידע הארגוני המוכר לי'. אל תנסה לנחש".

- **דרישה לציטוט מקורות:** כדי לייצר אמון, יש להנחות את הסוכן לציין בסוף כל תשובה את המקור המדויק שעליו הסתמך (למשל "מתוך: מדריך מקצועי, פרק 4, סעיף 2.1").

מנגנוני בקרה והרשאות

בניית סוכני AI לפי מתודולוגיית RAG היא תהליך עוקב בעל חוקיות של "תגובת שרשרת". טעויות או בעיות באיכות הנתונים שמתרחשות בשלבים המוקדמים, למשל בחילוץ הנתונים, מתגלגלות ומתעצמות ככל שמתקדמים לשלבים הבאים. פגמים אלו יובילו בהכרח לפגיעה חמורה באמינות הסוכן ול"הזיות" בשלב יצירת התשובה הסופית למשתמש. על כן הארגון חייב ליזום ניהול ובקרה דינמיים לאורך כל השלבים, כדי להבטיח מערכת איכותית ואמינה שניתן לקבל באמצעותה החלטות²⁴.

היזומה לניהול הנתונים דורשת הצבת שומרי סף ברורים לאורך כל התהליך:

- **בקרת אבטחת המידע:** יש לחשוף את המידע אך ורק בכפי העובדים שלהם יועדה המשימה. בשלב השליפה הווקטורית, חובה להטמיע מנגנון שבדק מי המשתמש ששואל את השאלה. סוכן ארגוני עשוי להכיל נהלים פומביים לצד טיוטות של דוחות ביקורת רגישים או נתוני משאבי אנוש. השאלתה צריכה לחפש רק בתוך הווקטורים המורשים לאותו משתמש ספציפי.
- **בקרת מומחי התוכן:** במיוחד בשלבי ההטמעה הראשונים (ובתהליכים רגישים גם לאורך השימוש השוטף), יש להגדיר תהליך שבו מומחי תוכן²⁵ סוקרים מדגם של שאלות עובדים ואת תשובות הסוכן עליהן. המטרה היא לזהות היכן כשל החיפוש הווקטורי (שלף מקטע לא נכון) או היכן יצר מודל השפה "הזיה" למרות ההנחיות ולתקן את אלגוריתם החיפוש או את אסטרטגיית חלוקת הנתונים (Chunking) בהתאם.
- **בקרת משתמשי הקצה:** יש לשלב מנגנון פשוט כמו דירוג איכות התשובה או כפתור "דווח על טעות" בממשק המשתמש. כך, כאשר עובד מזהה

24 Müller et al., Data Quality Challenges, לעיל ה"ש 16.

25 עובדים בעלי ידע מקצועי רחב או מומחיות בתחום מסוים.

26 משרד מבקר המדינה ונציב תלונות הציבור מדריך לביקורת המדינה (ללא ציון תאריך), <https://www.mevaker.gov.il/state-audit/roles-work/state-audit-guide>, כך כללי, הקדמה, דבר המנכ"ל.

הלשון לדייק באופן שוטף את הפרומפטים ובכך להבטיח שהסוכן יספק מענה מהימן לשאלות לשוניות ומרכבות ויבצע בקרה טכנית.

מהלך המחקר

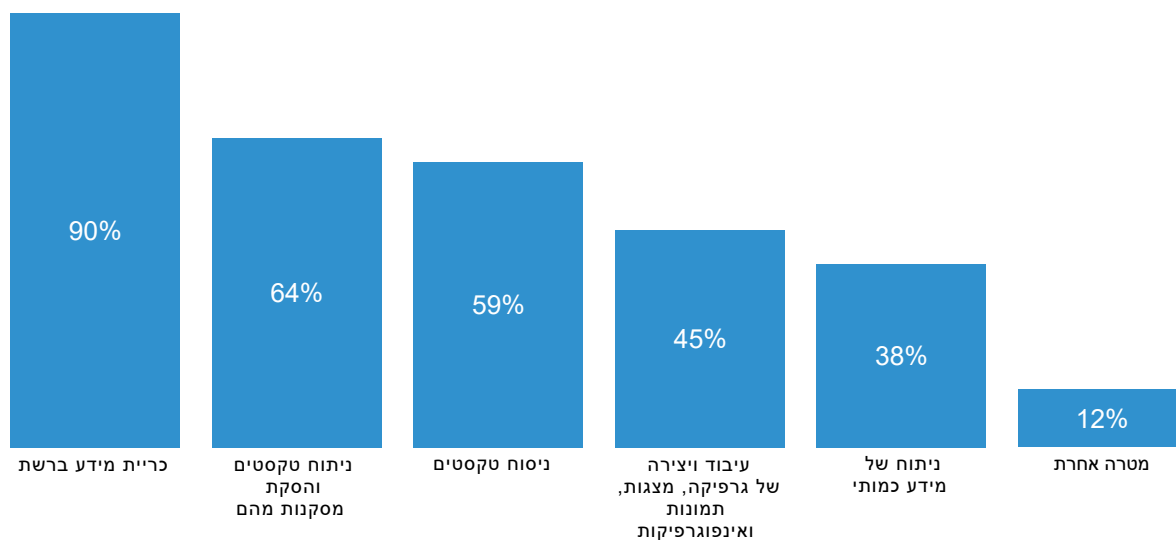
באפריל 2026 ערכנו סקר מקוון שהופנה אל 242 עובדים ומנהלים במשרד מבקר המדינה. מתוכם ענו על הסקר 76 עובדים ומנהלים. המשיבים נשאלו אם סוכני בינה מלאכותית ייעודיים המבוססים על הידע הארגוני של משרד מבקר המדינה יכולים לייעל את עבודת הביקורת, להבטיח אחידות מתודולוגית ולשפר את איכות הניסוח של הדוחות. הקפדנו על כך שאוכלוסיית הנשאלים תכלול אך ורק עובדים העוסקים בשגרה בכתיבת הדוחות ושותפים מלאים בתהליכי הבקרה עליהם. השאלון הורכב מארבעה חלקים עוקבים שנועדו לבחון את עמדות המשתתפים. החל מהיכרותם הכללית עם כלי בינה מלאכותית ועד להערכתם את הסוכנים הייעודיים שפותחו במשרד.

בשלב הראשון בשאלון נבחנו רמות ההיכרות והשימוש הנוכחי של הנשאלים בכלי בינה מלאכותית גנרטיבית (GenAI). הנשאלים נשאלו על הכלים שבהם הם משתמשים כיום, תדירות השימוש בהם והמידה שבה הם חשים כי כלים אלה תורמים לקיצור זמני העבודה ולשיפור איכות הדוחות. בשלב השני נבחנו עמדותיהם כלפי הפוטנציאל הגלום בשילוב מערכות בינה מלאכותית בתהליכי עבודת הביקורת.

רחב של דוחות ביקורת המדינה שפורסמו בשנים האחרונות, המשמשים דוגמאות חיות ליישום נכון ומדויק של כללי המדריך על שלל היבטיו.

2. **סוכן עריכת לשון** - הסוכן נועד לסייע לעובדי משרד מבקר המדינה בטיוב הניסוח, הדיוק הלשוני והאחידות הסגנונית של מסמכי המשרד. המטרה המרכזית בפיתוחו היא להנגיש את כללי העריכה המשרדיים לכלל הדרגים העוסקים בכתיבה ובכך להבטיח שדוחות הביקורת ובירור התלונות ינוסחו ברמה המקצועית והאיכותית ביותר, בהיצמדות קפדנית לשפה המוסדית הייחודית של המשרד. כדי לסייע לעובדי המשרד באופן אפקטיבי, הסוכן ניזון מתשתית מקצועית רחבה המורכבת משני נדבכים עיקריים: הנדבך הראשון הוא אוגדן עריכת הלשון, שהוא "ספר הכללים" המחייב של אגף התמיכה המקצועית. ממקור זה שואב הסוכן את הנחיות העבודה המפורטות, הכוללות פרקים מרכזיים כגון "תיקוני לשון ושימושי מילים", "מילון הכתיב המשרדי", "רכיבי דוח הביקורת" וכללי המבנה הצורני של הדוח. הנדבך השני שעליו נשען הסוכן הוא תהליך של מומחיות תוכן ושיפור מתמיד. הסוכן פותח ככלי דינמי שנועד לייעל את תהליכי העבודה מול יחידות הקצה. לאחר שלב הקמתו שותף הסוכן עם קבוצת מיקוד של עובדים ומנהלים לצורך התנסות ראשונית וניתנו הרשאות ניהול לעורכי הלשון המקצועיים ביחידה המשרדית לעריכת לשון. גישה זו מאפשרת לעורכי

תרשים ב: המטרות שעבורן הנשאלים משתמשים בכלי בינה מלאכותית



מהתרחשים ניכר כי המטרה המרכזית והשכיחה ביותר של שימוש ב-AI לצורך עבודת הביקורת היא כריית מידע ברשת, ו-90% מהנשאלים דיווחו שהם משתמשים בכלי בינה מלאכותית למטרה זו. נתון זה ממחיש את ההישענות ההולכת וגוברת על כלי בינה מלאכותית כמנועי חיפוש ואיסוף נתונים מתקדמים ויעילים. מייד לאחר מכן בולט שיעור השימוש הגבוה ביכולות עיבוד שפה טבעית: 64% מהמשתמשים דיווחו כי הם משתמשים בכלי AI לניתוח טקסטים והסקת מסקנות מהם ו-59% דיווחו כי הם משתמשים בכלים אלה לניסוח טקסטים. נתונים אלו מראים כי כלי AI הם תשתית קריטית לייעול עבודתם של אנשי מקצוע הנדרשים לעבד ולייצר כמויות גדולות של מידע טקסטואלי מורכב. לצד השימוש הנרחב בכלי AI לניתוח טקסטים וכריית מידע, הנתונים מצביעים גם על שימושים ויזואליים ואנליטיים נרחבים: 45% מהנשאלים דיווחו שהם משתמשים ב-AI לעיבוד ויצירה של גרפיקה, מצגות, תמונות ואינפורמיקות - נתון המעיד על חדירת הטכנולוגיה גם לתחומי העיצוב וההנגשה החזותית. 38% מהמשתמשים דיווחו על שימוש לצורך ניתוח של מידע כמותי כדי להפיק ממנו תובנות, מה שמדגיש את ערכה של הבינה המלאכותית ככלי תומך החלטות המבוסס על נתונים מספריים. לצד אלו, 12% מהנשאלים דיווחו על שימוש למטרות אחרות. לסיכום, התפלגות זו ממחישה כי עיקר הערך המוסף המופק כיום ממערכות בינה מלאכותית הוא ייעול של תהליכי איסוף המידע ושיפור הבנת הנקרא והניסוח. יכולות אלה הן אבן יסוד בכל מטלת ביקורת. בתרשים 3 מתוארת התפלגות המענה על השאלה "כיצד לדעתך תשפיע כניסתה של הבינה המלאכותית על אופי עבודת הביקורת במשרד בשנים הקרובות".

מהתרחשים עולה כי רוב מכריע של 75% מהמשיבים סבורים כי עבודת הביקורת תהפוך למעמיקה ואנליטית יותר בעקבות כניסת הבינה המלאכותית. נתון זה מתיישב בין היתר עם השאיפה המקצועית להעברת מרכז הכובד של המבקר מביצוע של ביקורת ציית לביקורת ביצועים מבוססת נתונים ומפעולות טכניות של איסוף נתונים לניתוח ביקורתי ומורכב. 72% מהמשיבים צופים כי תתפתח דרישה למומנויות חדשות בקרב המבקרים, מה שמעיד על הבנה כי הטמעת הכלים מחייבת התאמה של ההון האנושי והכשרה מקצועית רציפה.

עוד חשוב לציין שהציפיות בנוגע לגידול כמותי בתוצרי הביקורת נמוכות בהרבה. רק 22% מהמשיבים סבורים כי המשרד או אגפי הביקורת יוכלו להפיק מספר רב יותר של דוחות באותו פרק זמן. נתון זה עשוי להצביע על כך שהמבקרים צופים שהזמן שייחסך בזכות האוטומציה יושקע בשיפור העומק והאיכות של הביקורת הקיימת ולא דווקא בהגדלת היקף המטלות. לבסוף, התרחשים מפרך במידה רבה את החשש מהחלפת כוח האדם האנושי בטווח הקצר - רק 12% מהמשיבים סבורים כי השינוי

בשלב הבא ניתנה לנשאלים סקירה תמציתית של שני סוכני הבינה המלאכותית הייעודיים שפותחו במשרד: סוכן המדריך לביקורת וסוכן עריכת הלשון. הוסברו מטרותיהם של הסוכנים, מקורות הידע הארגוני שעליהם הם מבוססים והאופן שבו הם מיועדים להשתלב בתהליך העבודה. סוכנים אלו טרם הוטמעו באופן מלא בעבודה המשרדית, ולכן כדי לאפשר הערכה מבוססת ומעשית, שלב זה כלל סימולציה שבמסגרתה הוצגו למשתתפים דוגמאות מפורטות מתוך שיחות אותנטיות שנערכו עם הסוכנים. דוגמאות אלו כללו שאילתות מקצועיות ולשוניות מורכבות לצד התשובות וההנחיות שהפיקו הסוכנים בתגובה, לרבות ההפניות למקורות הידע הארגוניים הרלוונטיים (כגון סעיפים ספציפיים במדריך לביקורת או באוגדן עריכת הלשון). לאחר מכן התבקשו מנהלי הביקורת להעריך את איכות התשובות שהוצגו להם בסימולציה, את רמת הדיוק שלהן ואת מידת התאמתן לסטנדרט המקצועי והלשוני המשרדי.

לבסוף דירגו המשתתפים את מידת נכונותם לשלב כלים אלו בעבודתם השוטפת בעתיד ואת ציפיותיהם בעניין אופן השפעתם של הסוכנים על התייעלות עבודת הביקורת ואיכותה. ניתוח המענים בוצע בצורת שקלול (weighting) במטרה להתאים את התפלגות המדגם להתפלגות האוכלוסייה בפועל.

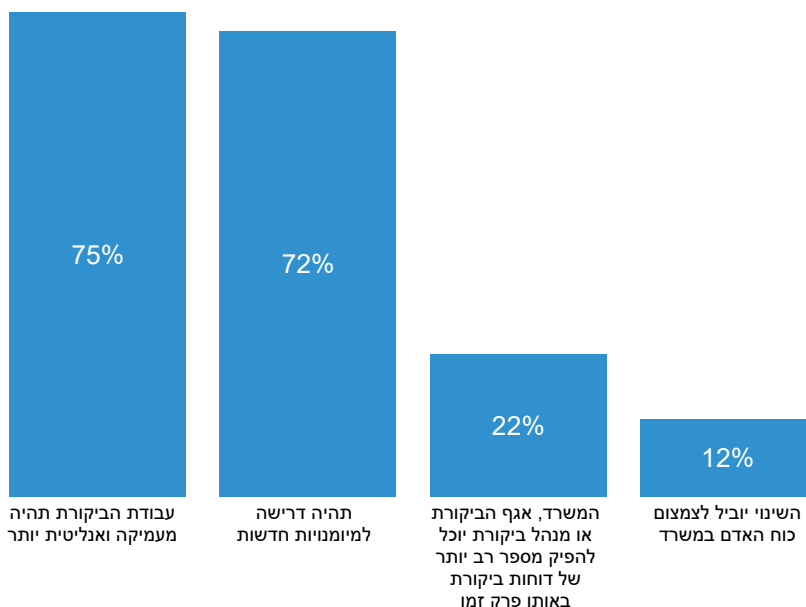
תוצאות

הסקר הועבר ל-242 עובדים ומנהלים, אשר חולקו לקבוצות על פי תפקידיהם והוותק שלהם, ו-76 מתוכם השיבו עליו. שיעורי ההיענות לא היו אחידים בין הקבוצות השונות, וכדי לתקן הטיה זו (non-response bias) חושבו משקולות דגימה לכל משיב, כך שהתרומה היחסית של כל קבוצה במדגם תשקף את חלקה באוכלוסייה הכוללת. המשקולות חושבו על בסיס היחס בין שיעור הקבוצה באוכלוסייה לבין שיעורה במדגם. לאחר החלת המשקולות, התקבלו אומדנים משוקללים המייצגים טוב יותר את האוכלוסייה כולה.

כבר מהמענה הראשוני לשאלון עולה ההשפעה הרבה של השימוש בכלי בינה מלאכותית. 86% מהנשאלים ציינו כי הם משתמשים בכלי בינה מלאכותית. יחד עם זאת, יש לציין כי התדירות של השימוש לצרכים פרטיים גדולה מתדירות השימוש לצורכי עבודה (79% ו-38% בהתאמה דיווחו על שימוש שבועי). בתרשים 2 מוצגת התפלגות השימוש של המשיבים לסקר בכלי בינה מלאכותית לצורך עבודת הביקורת.

מהתרחשים עולה תמונה ברורה של האופן שבו הבינה המלאכותית משתלבת בתהליכי הביקורת. הנתונים רלוונטיים במיוחד כאשר מנתחים את פוטנציאל ההטמעה של סוכני בינה מלאכותית בסביבות עתירות ידע וטקסט.

תרשים ג: הערכת ההשפעה של כניסת הבינה המלאכותית על אופי עבודת הביקורת



שיעור ההתנסויות שהוגדרו כשימושיות במידה רבה היה 47%, והתוצאות הפרטניות בדוגמאות השונות נעו בטווח שבין 53% ל-30%. ממציא זה ממחיש כי המודלים אומנם מצליחים להפיק מידע רהוט ומקצועי, אך התאמתו הישירה והמיידיית למורכבות תהליכי העבודה הארגוניים השוטפים היא עדיין אתגר. בבדיקת כל התשובות ניכר הבדל מובהק בין מנהלי הביקורת לבין הסגנים ומנהלי האגפים. מנהלי הביקורת דירגו גם את איכות התשובות של הסוכנים וגם את הפרקטיות שלהם בדירוג גבוה יותר ממנהלי האגפים והסגנים.

סיכום ומסקנות

מחקר זה בחן את שילובם של סוכני בינה מלאכותית המבוססים על ידע פנים-ארגוני במשרד מבקר המדינה, במטרה לייעל את עבודת הביקורת, להבטיח אחידות מתודולוגית ולשפר את איכות הניסוח של דוחות הביקורת. מן הממצאים עולה כי הנשאלים מרגישים שכניסת הבינה המלאכותית צפויה לחולל טרנספורמציה משמעותית באופי מקצוע ביקורת המדינה. רוב מכריע של המשיבים סבורים כי העבודה תהפוך לאנליטית ומעמיקה יותר וכי יידרש פיתוח מיומנויות חדשות. בניגוד להנחות שכיחות, העובדים והמנהלים אינם צופים צמצום בכוח האדם בטווח הקצר או גידול כמותי במספר הדוחות, אלא בעיקר הפניה של

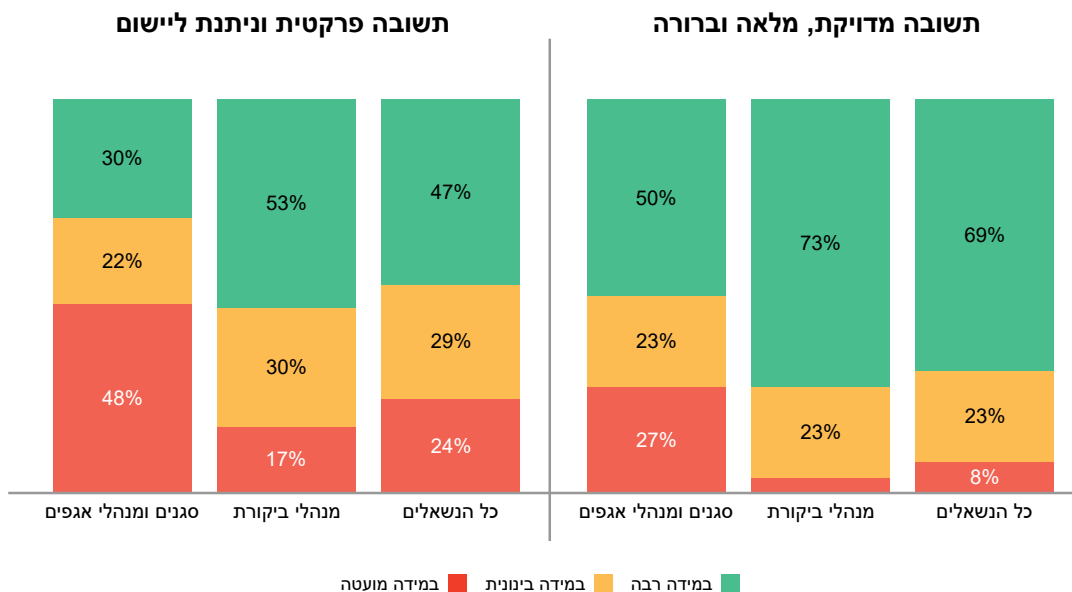
הטכנולוגי יוביל לצמצום כוח האדם במשרד. הממצאים מראים שהבינה המלאכותית נתפסת כזרז לטרנספורמציה מקצועית שיחזק את הערך המוסף של המבקרים באמצעות העמקת היכולות האנליטיות, וכי קיימת הכרה בכך שהמפתח להצלחה טמון ברכישת מיומנויות עבודה חדשות ומותאמות לעידן הדיגיטלי.

בתרשים 4 מתוארת הערכתם של הנשאלים את איכותם ומידת יישומיותם של סוכני הבינה המלאכותית שנבחנו.

המבחנים הסטטיסטיים שנערכו (p-value של 0.5 למדד האיכות ו-0.9 למדד היישומיות) מצביעים על כך שאין הבדל מובהק סטטיסטית בין ביצועי הדוגמאות השונות שהוצגו לנשאלים. הנתונים בגרף מציגים את הדירוג הממוצע שהעניקו המשיבים לתוצרי הסוכנים שהוצגו להם במסגרת השאלון. ממדד האיכות, הבוחן עד כמה התשובות שהתקבלו היו מדויקות, מלאות וברורות, עולה תמונה חיובית - 69% מכלל המשיבים דירגו את התוצרים כאיכותיים במידה רבה. בחינה פרטנית של הדוגמאות שהוצגו לנשאלים מתוך שיחות אותנטיות שנערכו עם הסוכנים מצביעה על שיעורי שביעות רצון גבוהים שנעו בין 50% ל-73%, נתון המעיד על יכולות עיבוד ושפה מתקדמות.

עם זאת, בבחינת מדד הפרקטיות, הבדוק עד כמה התשובות היו שימושיות וניתנות ליישום ישיר הלכה למעשה בעבודה, ניכר פער יישומי מסוים. בראייה כוללת,

תרשים ד: הערכת איכות התוצרים של סוכני הבינה המלאכותית



דרגי הניהול. מנהלי ביקורת מהשטח דירגו את התוצרים כאיכותיים ופרקטיים יותר ממנהלי האגפים וסגניהם. לסיכום, סוכני בינה מלאכותית מציעים פוטנציאל רב לצמצום טעויות אנוש, ייעול משימות שגרתיות ומתן פתרון לבעיית "התמחות היתר" באמצעות שיתוף ידע ארגוני קולקטיבי. עם זאת, הצלחתם של סוכנים אלה תלויה לא רק בפיתוח הטכנולוגי אלא בנכונות ההנהלה והעובדים לאמץ אותם כחלק בלתי נפרד מהתרבות הארגונית, לשתף עימם ידע כדי להגדיל את מאגר המידע ולהמשיך להתנסות בהם כדי לדייק את השימוש. המפתח להשגת יתרון מקצועי בר-קיימא טמון בצמצום הפער היישומי של המודלים ובהשקעה מתמדת בלמידה לאורך החיים²⁷ כדי להבטיח שעובדי הביקורת ירכשו את המיומנויות הנדרשות לעבודה אפקטיבית, אחראית וביקורתית לצד הטכנולוגיה החדשה, וזאת כדי שיפותרו עוד אמצעיים מתקדמים שישרתו את הביקורת.

המשאבים שיתכנו לשיפור האיכות והעומק של התוצרים. כבר כיום השימוש המרכזי בכלים אלה במשרד מתמקד בכריית מידע, בעיבוד שפה טבעית ובניסוח טקסטים - יכולות שהן אבן יסוד בעבודת הביקורת עתירת הטקסט. עם זאת, בחינת התוצרים של שני הסוכנים הייעודיים שפותרו במשרד - סוכן המדריך לביקורת וסוכן עריכת הלשון - חושפת תמונה מורכבת המוגדרת כ"פער יישומי". מרבית המשיבים העריכו את תשובות הסוכנים כמדויקות, מלאות וברורות במידה רבה, אך פחות ממחציתם מצאו אותן פרקטיות וניתנות ליישום ישיר הלכה למעשה בעבודתם. פער זה מדגיש כי אף שהמודלים מצליחים להפיק מידע מקצועי ורהוט, התאמת התוצרים למורכבות ולדקויות של תהליכי העבודה השוטפים בארגון היא עדיין אתגר. יש לסייג ולציין כי סוכנים אלו טרם הוטמעו בעבודה המשרדית, ולכן ייתכן שההערכות ישתנו לאחר התנסות ושימוש רציפים לאורך זמן. כמו כן, ניכר הבדל מובהק בין



השלם גדול מסך חלקיו: מטא-ביקורת בשלטון המקומי בעידן הבינה המלאכותית

אילת רייכמן-עמית

עו"ד אילת רייכמן-עמית היא סגנית מנהל האגף לביקורת גופי השלטון המקומי - החוף, השפלה והנגב בחטיבה לביקורת השלטון המקומי שבמשרד מבקר המדינה ונציב תלונות הציבור.

המאמר בקצרה

- המטא-ביקורת (Meta-Audit) היא פרדיגמה המשלבת בחינה מתכללת של ממצאי ביקורת קודמים עם ניתוח רפלקטיבי של תהליך הביקורת עצמו.
- התהליך מאפשר חילוץ "תובנות-על" (Meta-Insights) מצרפיות שלא ניתן לזהותן מקריאה של דוחות ביקורת בודדים.
- מאפייני השלטון המקומי, ובהם "רעש נתונים" ומבנה דו-רובדי ומבוזר, הופכים אותו למתאים במיוחד ליישום תהליך זה.
- שילוב בינה מלאכותית מהווה מכפיל-כוח ומאפשר ניתוח מובנה של מאות דוחות ביקורת במהירות גבוהה - פעולה שאי אפשר ליישם באמצעים אנושיים בלבד.
- הצלחת הפרדיגמה בכלי בינה מלאכותית מחייבת שילוב של מסגרת עבודה זהירה ואחראית כדי להבטיח שהטכנולוגיה לא תהיה זו שמקבלת את ההחלטה, אלא תשמש רק כלי תומך בתהליך קבלתה.

מבוא | המסגרת התיאורטית: מטא-ביקורת וביקורת המדינה | מתיאוריה לפרקטיקה: היישום המעשי של מטא-ביקורת בכלי הבינה המלאכותית | סיכונים ואתגרים ביישום המטא-ביקורת בכלי בינה מלאכותית | סיכום

יוצגו מגבלות אתיות ומקצועיות ליישום המטא-ביקורת והשפעותיהן האפשריות על תחום ביקורת המדינה.

המסגרת התיאורטית: מטא-ביקורת וביקורת המדינה

יעדיה של ביקורת המדינה אומנם עשויים להשתנות לפי סוג הביקורת, אך לפי תקני הביקורת הבין-לאומיים, עקרונות היסוד לתרומתה הם זהים: הקניית ידע, מסקנות והמלצות עצמאיות ומבוססות לחיזוק האחריות, השקיפות ויעילות השימוש בכספי ציבור ולשיפור מתמיד בהתנהלות הגופים העומדים לביקורת.² מימוש העקרונות הללו במובנם הרחב עשוי לחייב זיהוי של דפוס ליקויים חוזר ובעיות שורש בהתנהלותם של הגופים העומדים לביקורת וכן זיהוי ליקויים מערכתיים וכשלים במנגנוני האסדרה.³ לשם כך עשוי להידרש תכלול של ממצאים מכמה דוחות ביקורת שנעשו על אותם הגופים המבוקרים בפרקי זמן שונים, על אותם הגופים המבוקרים בנושאים שונים ואף על גופים מבוקרים שונים באותם הנושאים.⁴ מימוש העקרונות הללו עשוי לחייב גם בחינה רפלקטיבית של תהליך הביקורת עצמו, כדי לזהות פערים בנושאים או בגופים שנבדקו ולהבטיח להבא כיסוי מיטבי בביקורת של נושאי הביקורת והגופים העומדים לביקורת.

הפעולות המתוארות הן חלק מפרדיגמה חדשה שתקרא במאמר זה "מטא-ביקורת" - תהליך המשלב בחינה מתכללת של ממצאים והמלצות מתוך דוחות ביקורת קודמים ובחינה רפלקטיבית של תהליך הביקורת עצמו, בהסתכלות רב-שנתית ורב-נושאית; זאת במטרה לחלץ תובנות מצרפיות שאי אפשר לזהותן בהכרח מתוך קריאה

משרד מבקר המדינה מפרסם בכל שנה עשרות דוחות ביקורת במגוון רחב של נושאים. מתכונת הפרסום של הדוחות מקשה לבחון בחינה רוחבית את נושאי הביקורת שנבדקו בפרק זמן מסוים, ובפרט על פני כמה שנים, וכן לבחון בחינה רוחבית את הגופים שבהם נעשתה הביקורת ואת הממצאים וההמלצות בעניינם.¹ חסרוננו של מבט רוחבי כזה ניכר במיוחד בדוחות הביקורת בשלטון המקומי, שמתאפיין במבנה מבוזר ומורכב ושבידקתו מקיפה מאות גופים מבוקרים.

מאמר זה מבקש להציע פרדיגמה חדשה בתחום ביקורת המדינה - המטא-ביקורת (Meta-Audit) - ולבחון את יישומה המעשי באמצעות כלי בינה מלאכותית על דוחות ביקורת המדינה בשלטון המקומי בישראל. המאמר מבקש לבחון אם שימוש בכלי בינה מלאכותית לניתוח מתכלל של ממצאי מאות דוחות ביקורת במבט רטרוספקטיבי ורפלקטיבי יוכל להיות כלי אסטרטגי לייעול התכנון של עבודת הביקורת ולטיובה; זאת דרך בחינת המודל של NotebookLM כמקרה בוחן, הצגת מתודולוגיה לתחקור דוחות עבר ותיקור המשלב בקרה אנושית (Human-in-the-Loop).

בפרק הראשון של המאמר יוצגו מושג המטא-ביקורת ושורשיו הרעיוניים. כמו כן תוצג המטא-ביקורת כשלב נוסף וחיוני בהתפתחותה של ביקורת המדינה, וייסקרו המאפיינים הייחודיים של ביקורת המדינה בשלטון המקומי ההופכים אותה למתאימה ביותר לבחינת יישומה של הפרדיגמה המוצעת. הפרק השני ידון בתוצאותיה של בדיקת היתכנות ראשונית (Proof of Concept) של יישום המטא-ביקורת בכלי בינה מלאכותית. בפרק השלישי

1 הדוחות נכתבים ומתפרסמים כקובצי טקסט נפרדים. משכך, ניתוח רוחבי מעמיק של נושאי הביקורת או הגופים שבהם נעשתה הביקורת מחייב פתיחה של כל קובץ בנפרד ואיתור ממצאים רלוונטיים מתוכו.

2 תקן ISSAI 100 "עקרונות יסוד של ביקורת המזר הציבורי", ס' 20. עקרונות היסוד הללו עולים בקנה אחד עם התפקידים ועם הסמכויות שהוקנו בדיון למבקר המדינה בישראל: לבדוק אם פעולותיהם של הגופים הנתונים לביקורת נעשות בהתאם לחוק, למינהל התקין ולטוהר המידות ועל פי עקרונות של יעילות וחיסכון, וכן לבדוק כל עניין אחר שיראה בו צורך. ראו חוק מבקר המדינה, התש"ח-1958 [נוסח משולב], ס' 10; חוק יסוד: מבקר המדינה, ס' 2.

3 משמעו של כשל אסדרה (Regulatory Failure) הוא כי מדיניות האסדרה שנקבעה אינה המדיניות המיטבית, לדוגמה בשל אסדרה חלקית או חסרה, בשל אי-בהירות של הוראות רגולטוריות או סתירה ביניהן או בשל היעדרם של כלי אכיפה מספקים. ראו והשוו יובל רויטמן "הרפורמה הרגולטורית: בין הגלוי לסמוי" 425 (מסדרת **משפט, חברה ותרבות**, ישי בלנק, דוד לוי-פאור ורוני קרייטנר עורכים 2016); Barak Orbach, What is Government Failure, 30 (2012). Yale J. Reg. Online 1 (2012). (Systemic Failure) מתייחס לליקויים חוצי ארגון הקשורים במבנה ובמורכבות של המערכת, בתהליכי העבודה או בתרבות הארגונית, לדוגמה ריבוי גופים מעורבים והיעדר מנגנוני תיאום ביניהם או פערים בין הסמכות לאחריות של הגופים הללו. כשל אסדרה יכול להוביל לכשל מערכתי ולהפך, אם לדוגמה היעדרה של אסדרה או אסדרה חסרה מובילים לחוסר תיאום בין גופים או לפערים בתחומי הסמכויות או האחריות שלהם, או שהם נגרמים מחוסר תיאום או מפערים כאלה.

4 אף על פי שתקני הביקורת הבין-לאומיים אינם מתייחסים מפורשות לפעולה של תכלול ממצאים מכמה דוחות ביקורת המדינה, קיימת בתקני הביקורת התייחסות לפעולה של תכלול ממצאים מתוך דוחות שעשו גופים אחרים לשם הערכת מדיניות ציבורית - פעולה המכונה בתקנים הבין-לאומיים "מטא-הערכה". פעולה זו נועדה לשמש "לקביעת התועלת, התקינות, השיטיות והדיוק של הערכה אחת או של הערכות רבות". ראו תקן ISSAI 9020 "מדיריך להערכת מדיניות ציבורית", ס' 3.2.

וחוזקותיהם, ובשונה מהמטא-אנליזה אפשר ליישם אותה גם על מחקר הערכה בודד⁹.

הפרדיגמה המוצעת במאמר זה - המטא-ביקורת - מבקשת לשלב בין עולמות המטא-אנליזה והמטא-הערכה עם ההתאמות המתחייבות לעולם ביקורת המדינה. מוצע לבסס את הפרדיגמה החדשה, בדומה למטא-אנליזה, על בחינה מתכללת של ממצאים והמלצות מדוחות עבר של משרד מבקר המדינה, גם אם ללא ניתוח סטטיסטי שיטתי, במטרה לזהות דפוס ליקויים חוזר בהתנהלותם של הגופים העומדים לביקורת ובעיות שורש בהתנהלותם וכן ליקויים מערכתיים וכשלים במנגנוני האסדרה שלהם. בדומה למטא-הערכה, מוצע כי המטא-ביקורת תשמש לא רק להערכה של ממצאי הביקורת, אלא גם להערכה של תהליך הביקורת וכך תסייע בנייתו מגמות בביקורת המדינה לאורך השנים ובזיהוי תחומים שלא נכללו בביקורת או גופים מבוקרים שלא נבדקו זמן רב. בהקבלה לתהליך המטא-אנליזה, תוצאות תהליך המטא-ביקורת המוצע עשויות לשמש גם לבחינת נחיצותן של ביקורות חדשות ולשמש בסיס לתכנון; הן יכולות לשמש גם להבנת ההקשר של ביקורות חדשות, כהמשך לביקורות קודמות שנעשו על אותם גופים מבוקרים או בנושאים דומים¹⁰; והן יכולות לשמש גם לבחינת השפעתן של ביקורות קודמות על התנהלותם של הגופים המבוקרים¹¹. מאפייני מודל המטא-ביקורת המוצע מוצגים ב שלהלן.

של דוחות בודדים והפיכתן לכלי אסטרטגי לתכנון עבודת הביקורת ולטיובת. תובנות אלו ייקראו "תובנות-על" (Meta-Insights)⁵.

התשתית הרעיונית של המטא-ביקורת

שורשיו הרעיוניים של מושג המטא-ביקורת הם בעולמות המטא-אנליזה (Meta-Analysis) והמטא-הערכה (Meta-Evaluation)⁶. מטא-אנליזה היא תהליך סינתזה (תכלול) של ממצאים מצטברים מתוך מחקרים מדעיים נפרדים בנושא מסוים במטרה להגיע למסקנה אחת, קוהרנטית ותקפה מדעית, באותו הנושא, לעיתים כחלק מסקירה מדעית שיטתית (Systematic Review)⁷. השימוש הנפוץ ביותר של המטא-אנליזה הוא כבסיס לקביעת מדיניות בתחומי המחקר השונים, אולם היא עשויה למלא תפקיד משמעותי גם בתהליך המחקרי עצמו, וממצאיה עשויים לסייע בבחינת עצם הצורך בעריכתו של מחקר חדש, ולסייע גם בתכנונו, אם נמצא שהמחקר נחוץ. מטא-אנליזה יכולה לסייע גם בהמחשת הערך המוסף של המחקר החדש ובהבנת ההקשר הרחב שלו בתוך שדה המחקר הקיים, ובכך למנוע מצב הקרוי במחקר המדעי "Islands in Search of Continents" - נתק בין מחקרים חדשים ("איים") לשדה המחקר הקיים ("יבשות")⁸. לעומת זאת, מטא-הערכה משמשת לבחינה איכותנית של מחקרי הערכה קודמים ושל תוצריהם, לשם הערכת חולשותיהם

5 לפי מילון "רב-מילים", משמעה של המילה תובנה הוא "הבנת המשמעות או המהות האמיתית או העמוקה של דבר, בעיקר מתוך תפיסה אינטואיטיבית, לעיתים כתאומית".

6 מחקר הערכה הוא מחקר המתמקד בבחינת האיכות, היעילות וההשפעה של תוכניות חברתיות בתחום מדיניות מסוים. ראו Peter H. Rossi, Mark W. Lipsey & Howard E. Freeman, Evaluation: A Systematic Approach (7th ed. 2004).

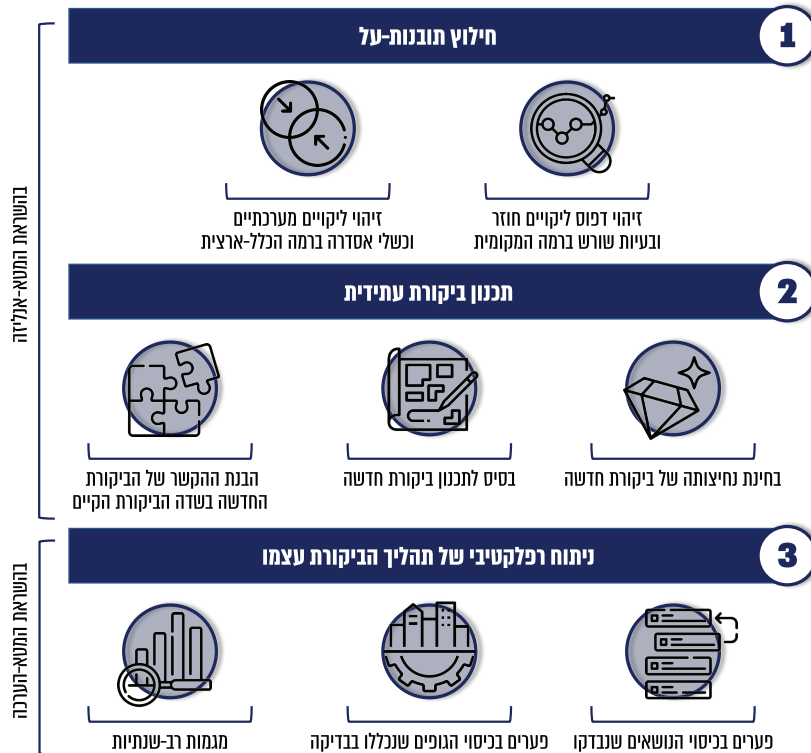
7 שלב מוקדם יותר בהתפתחות המטא-אנליזה הוא שיטת הסקירה הנרטיבית (Narrative Review) שכללה סיכום של ממצאי מחקרים והסקת מסקנה אחת מתכללת מהם.

8 Michael Borenstein et al., Introduction to Meta-Analysis xxi-xxii (2009), ובתרגום לעברית "איים בחיפוש אחר יבשות", טבעו בשנת 1998 החוקרים איאן צ'למס ומייקל קלארק. ראו Michael Clarke & Iain Chalmers, Discussion Sections in Reports of Controlled Trials Published in General Medical Journals: Islands in Search of Continents?, 3 JAMA 280 (1998).

9 Daniel L. Stufflebeam, George F. Madaus & Thomas Kellaghan (eds.), Evaluation Models 458 (2000); Michael Scriven, Meta-Evaluation Revisited, 6(11) Journal of MultiDisciplinary Evaluation iii (2009).

10 ההקשר של דוח ביקורת מסוים בשדה הביקורת מובא כיום באופן שיטתי בעיקר בדוחות ביקורת מעקב, ולעיתים ישנה התייחסות לדוחות ביקורת קודמים בנושאים דומים, אך לא באופן שיטתי. ההתייחסות מובאת בפרק שכותרתו "פעולות הביקורת", שמופיע בכל דוח ביקורת המתפרסם, ושכולל פירוט על מטרת הביקורת ונושאייה והגופים שנבדקו, ולעיתים גם מידע נוסף על תהליך הביקורת, כלי הביקורת ומתודולוגיית הביקורת.

11 לדוגמה, באמצעות ניתוח השוואתי של ממצאי ביקורת על גוף מבוקר מסוים בנושאי ביקורת שונים העשויים להצביע על התרבות הארגונית של אותו הגוף. ניתן לשלב בתהליך המוצע בחינה מתכללת גם של מועילות הביקורת, במובן של הנעה לפעולה של הגוף המבוקר ותיקון הליקויים על ידו, החסמים העומדים בפניו ליישום המלצות הביקורת וההשפעה הציבורית של תיקון הליקויים או של אי-תיקונם. היבטים הנוגעים למועילות הביקורת מוסדרים בחלקם בתקני הביקורת הבין-לאומיים וחלקם נבחנים כבר כיום, ללא הסתייגות בכלי בינה מלאכותית. שילובם בתהליך מחייב ניתוח של מידע ונתונים החיצוניים לדוחות הביקורת, ומשכך הם לא יידונו במאמר זה המתייחס לאפשרות ליישם מטא-ביקורת בהיבטים שטרם נבחנו, תוך ניתוח מתכלל של דוחות הביקורת בלבד.



דוחות ביקורת בתחום מדיניות מסוים במטרה לזהות אתגרים וסיכונים אסטרטגיים באותו תחום, בלי להוסיף על הממצאים או ההמלצות שנכללו בדוחות הביקורת המקוריים¹³. מודל דומה, ה-Insight Report ("דוח תובנות"), משמש בעולמות הביקורת הפנימיים¹⁴. מודל אחר של בית המשפט האירופי לביקורת הוא ה-Meta-Audit Special Report, המבוסס על תכלול של תובנות מתוך דוחות עבר בנושא מסוים, במטרה לזהות פערים בביקורת שנעשו

מודלים של מטא-ביקורת במבט השוואתי

מוסדות ביקורת המדינה בעולם אימצו מודלים שונים לתכלול ממצאים מכמה דוחות ביקורת, במטרה להגביר ולהעמיק את תרומתם הציבורית. כך לדוגמה, מודל ה-Lessons Learned הבריטי¹² ומודל ה-Landscape Review של בית המשפט האירופי לביקורת (European Court of Auditors; ECA) מבוססים על סקירה מתכללת של כמה

12 לדוגמה: National Audit Office (NAO), Audit Insights: Lessons and findings from the NAO's Financial Audits in 2024-25 (2026); National Audit Office (NAO), Lessons learned: The government's use of external consultants (2025).

13 European Court of Auditors, Gaps, overlaps and challenges: A landscape review of EU accountability and public audit arrangements (2014); European Court of Auditors, Landscape Review: EU action on energy and climate change (2017); European Court of Auditors, Landscape Review: Putting EU law into practice: The European Commission's oversight responsibilities under Article 17(1) of the Treaty on European Union (2018).

14 לדוגמה, דוחות "ריכוז ממצאים" שמפרסמת רשות ניירות ערך בנושאים שונים, שמטרתם לתכלול ממצאים עיקריים שעלו בבדיקות או בביקורת פרטיות, בלי לכלול הערות המופנות ישירות לגופים שנבדקו. ראו בועז ענר "הפקת תובנות ממקבץ דוחות ביקורת פנימית וציבורית" **כתב העת של לשכת המבקרים הפנימיים IOA ישראל, גיליון 05 - צעד אחד קדימה** (אפריל 2017).

הממשל הפדרלי¹⁶. עיקרי ההבדלים בין המודלים הקיימים המשמשים לתכלול דוחות ביקורת מובאים ב שלהלן. המודלים הקיימים לתכלול של תובנות מתוך דוחות קודמים נועדו ככלל לזהות דפוס ליקויים חוזר בגוף מבוקר מסוים או כשלי אסדרה בתחום פעילות מסוים. חלקם משמשים באופן מוגבל לחילוץ של תובנות חדשות מתוך דוחות העבר, לדוגמה על מגמות רב-שנתיות או על קשרים בין-נושאיים. חלק מהמודלים הקיימים כוללים בחינה

ולתכנן ביקורות עתידיות בנושא¹⁵. דוגמה נוספת היא המודל האמריקאי של ה-Hight-Risk List שמפרסמת לשכת המבקר הפדרלית של ארצות הברית מאז שנת 1990, ועם כינונו של כל קונגרס חדש. מודל זה מבוסס על ניתוח מתכלל של דוחות ביקורת קודמים, והוא כולל סקירה של תחומי הביקורת המצויים בסיכון הגבוה ביותר לחוסר יעילות בניהול כספי ציבור או להונאה. מטרת הניתוח היא לשמש מפת דרכים לניהול סיכונים ברמת

לוח 1: מודלים קיימים לתכלול ממצאים מתוך כמה דוחות ביקורת

Meta-Audit Special Report	High-Risk List	Lessons Learned	Landscape Review	Insight Report	
					תכלול מידע מכמה דוחות ביקורת
					זיהוי דפוס ליקויים חוזר וכשלי אסדרה
					חילוץ תובנות חדשות וצפיות פני עתיד ⁽¹⁾
					ניתוח רפלקטיבי של תהליך הביקורת ⁽²⁾
					שימוש משולב פנים-ארגוני וחץ-ארגוני ⁽³⁾
					ניתוח רב-ממדי ⁽⁴⁾
					יחידת ניתוח משולבת מקומית וכלל-ארצית ⁽⁵⁾

(1) לדוגמה, לצורך תכנון ביקורות עתידיות או לצורך ניהול סיכונים אסטרטגי.
 (2) שינויים לאורך השנים בנושאים שנבדקו, בכיסוי של הגופים העומדים לביקורת ובמתודולוגיית הביקורת או הפערים בנושאים שנכללו בביקורת, וכן השפעתה של הביקורת על התנהלות הגופים המבוקרים.
 (3) מודל שקהל היעד שלו הוא לא רק הגופים העומדים לביקורת או מקבלי ההחלטות וקובעי המדיניות בנושאים הנבדקים בביקורת, אלא גם מוסד הביקורת ומנהלי הביקורת עצמם.
 (4) ניתוח של קשרים בין נושאי ביקורת שונים שנבדקו בגוף מבוקר מסוים או קשרים בין גופים מבוקרים שנבדקו באותו הנושא בביקורות נפרדות.
 (5) שילוב של תובנות ברמה המקומית (גוף מבוקר בודד) לצד תובנות ברמה הכלל-ארצית (כשלי אסדרה, כשלים מערכתיים וניתוח רפלקטיבי על תהליך הביקורת עצמו).

European Court of Auditors, EU pre-accession assistance for strengthening administrative capacity in the Western Balkans: A meta- 15 audit (2016).

16 מתוך אתר המרשתת של לשכת הביקורת הפדרלית של ארצות הברית. www.gao.gov/high-risk-list

רפלקטיבית של תהליך הביקורת עצמו, בחינה רב-ממדית המשלבת ניתוח של ממצאים על גופים שונים באותו הנושא או בחינה של ממצאים על נושאים שונים שמהם אפשר לחלץ תובנות חדשות. אך אף לא אחד מהמודלים הקיימים כולל את הרכיבים הללו יחדיו, כפי שמוצע לכלול במסגרת הפרדיגמה החדשה של המטא-ביקורת, כמפורט להלן.

מטא-ביקורת כשלב נוסף בהתפתחות ביקורת המדינה

לנוכח מספרם הרב של דוחות הביקורת שמפרסם משרד מבקר המדינה - עשרות דוחות מדי שנה בשנה, אופן הכנתם וכתביבתם ומתכונת פרסומם כקובצי טקסט נפרדים - אין זה פשוט לתכלל ממצאים מכמה דוחות ולחלץ מהם תובנות-על. הקושי הרב הכרוך בכך אף גובר על רקע התמורות שחלו עם השנים במאפייניה של ביקורת המדינה.

שינוי מרכזי אחד שחל במאפייניה של ביקורת המדינה נוגע לנושאים העומדים לביקורת. שינוי זה מתבטא במעבר מביקורת על החוקיות והסדירות של הניהול הכספי והחשבונאי של הגופים המבוקרים (ביקורת ציוד) לביקורת על היעילות, המועילות והחיסכון בהתנהלותם (ביקורת ביצועים)¹⁷. שינוי נוסף נוגע למידע הנבדק בביקורת ולכלים המשמשים לניתוח המידע ולעיבודו, שינוי המתאפיין במעבר מקריאה ועיבוד של מסמכים המתקבלים מהגופים המבוקרים (ביקורת מסורתית), לניתוח עצמאי של מסדי נתונים ושימוש בכלים טכנולוגיים חדשניים לשם כך (ביקורת חדשנית)¹⁸. על רקע זה חל שינוי משמעותי גם במתודולוגיית הביקורת, שינוי שהתגבר במיוחד בשנים האחרונות על רקע ההתפתחות הטכנולוגית וחזונו של המבקר הנוכחי, ומתאפיין במעבר מביקורת המתבצעת בגוף עיקרי אחד ומתמקדת בנושא אחד או בכמה נושאים (ביקורת נקודתית-אנכית) לביקורת המתבצעת בעת ובעונה

אחת בכמה גופים מבוקרים, שבכולם נבדק אותו הנושא באופן השוואתי (ביקורת אופקית-מערכתית)¹⁹.

התמורות שחלו במאפייניה של ביקורת המדינה הגבירו את מורכבותו של תהליך הביקורת, וגם תוצרי הביקורת - הממצאים וההמלצות המובאים בדוחות - נעשו מורכבים יותר²⁰. לא מן הנמנע כי תמורות אלו הביאו לכך שהדוחות נהיו ארוכים יותר. בכך טמון אתגר רב ליישום המטא-ביקורת, אך בה בעת יש בכך גם כדי להעיד ביתר שאת על החיוניות שביישומה ועל היותה שלב חיוני נוסף בהתפתחות ביקורת המדינה, להבטחת מועילותה ותרומתה הציבורית. התחומים שבהם חלו תמורות במאפייניה של ביקורת המדינה לאורך השנים מובאים ב שלהלן.

מאפייניה הייחודיים של ביקורת המדינה בשלטון המקומי בישראל מחייבים דיון נפרד בכל הקשור לאתגרים הכרוכים ביישומה של המטא-ביקורת בתחום זה. מאפיין בולט וייחודי של ביקורת המדינה בשלטון המקומי הוא "רעש נתונים" (Noise Data) מוגבר. הסיבה היא שתחום ביקורת זה משתרע על פני כ-1,500 גופים מבוקרים המתאפיינים בשונות רבה ובפריסה גיאוגרפית נרחבת²¹. מאפיין ייחודי נוסף של ביקורת המדינה בשלטון המקומי הוא **הדואליות המובנית**, שמקורה במבנה הדו-רובדי והמבוזר של הממשל בישראל²². מבנה זה מחייב להידרש בביקורת לא רק לבחינת הליקויים בהתנהלותן של הרשויות המקומיות בנושא מסוים, אלא גם לבחינת הכשלים באסדרה, בחלוקת הסמכויות או בהקצאת המשאבים, ומכאן שישנו הכרח בבחינה משולבת של השלטון המרכזי והמקומי גם יחד בדוח ביקורת אחד.

מאפייניה הייחודיים של ביקורת המדינה בשלטון המקומי ניכרים במיוחד בשנים האחרונות בשל ריבוי הביקורות המערכתיות בשלטון המקומי המתבצעות בעת ובעונה אחת בכמה גופים מבוקרים, ובכולם נבדק אותו הנושא

17 ראו אורנית צירקין "ביקורת ביצועים של ביקורת המדינה בישראל בנושאים חברתיים: מורכבות הביקורת ומבט השוואתי" 64 **מראות: עיונים בביקורת המדינה ובבירור תלונות הציבור** 108 (2024).

18 ראו האלה אבו חאלה ואריאל סוארי "ניתוח נתונים בביקורת מדינה: מביקורת מסורתית לביקורת חדשנית - ממה להיזהר?" 65 **מראות: עיונים בביקורת המדינה ובבירור תלונות הציבור** 7 (2025).

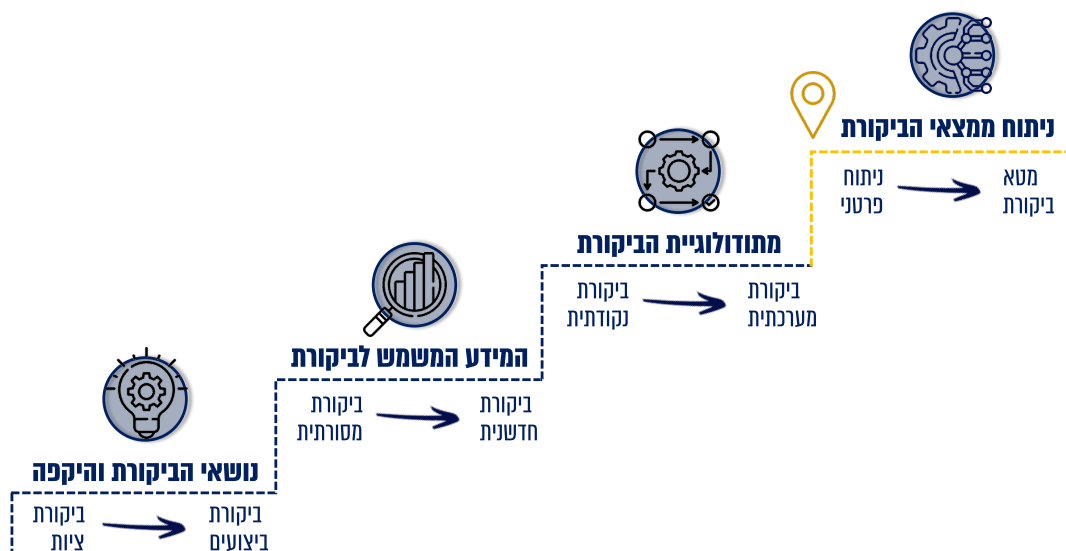
19 ראו ניסים מזרחי "ביקורת המדינה והביקורת הפנימית בישראל - היבטים השוואתיים" 61 **עיונים בביקורת המדינה** 135 (2008); אילת רייכמן-עמית "ללמוד לבחור נכון: מערכת לניהול ובחירה של רשויות מקומיות לביקורת (מנ"ר)" 65 **מראות: עיונים בביקורת המדינה ובבירור תלונות הציבור** 159 (2025).

20 לדוגמה, בביקורות מערכתיות נעשה מורכב הרבה יותר לזהות ממצא המתייחס לגוף מבוקר יחיד, שלא נכלל בביקורת העומק, מבין הממצאים המערכתיים המובאים בדוח.

21 בנוגע ל"רעש הנתונים" המאפיין את תחום ביקורת המדינה בשלטון המקומי, כבר עלתה בעבר טענה לגבי מועילות ביקורת זו מבחינת יכולת הכיסוי של כלל הגופים המבוקרים ותחומי פעילותם הרבים. ראו בנימין גייסט "לסוגיית האפקטיביות של ביקורת המדינה ברשויות המקומיות בישראל" **הביקורת ברשויות המקומיות** 87 (אשר פרסם עורך 1999).

22 נחום בן-אליא **הדור הרביעי - שלטון מקומי חדש לישראל** (2004); ישי בלנק "מקומו של המקומי: משפט השלטון המקומי, ביזור ואי-שוויון מרחבי בישראל" **משפטים** 197 (2) (התשס"ד).

תרשים ב: התמורות במאפייניה של ביקורת המדינה לאורך השנים



הזדמנות לממש את יכולותיה של המטא-ביקורת, באופן שאינו ברייטום בכלים אנושיים בלבד, ובכך להביא למלוא המימוש של עקרונות היסוד בעבודת ביקורת המדינה ולקידום היעילות והמועילות של המינהל הציבורי באופן המיטבי.

מתיאוריה לפרקטיקה: היישום המעשי של מטא-ביקורת בכלי הבינה המלאכותית

היישום המעשי של המטא-ביקורת בעולמות הבינה המלאכותית נשען על יכולות טכנולוגיות מתקדמות של עיבוד שפה טבעית (Natural Language Processing; NLP)²³. יכולות אלו מיושמות באמצעות מודלי שפה גדולים (Large Language Models; LLMs)²⁴ המאפשרים ניתוח ממוחשב ומהיר מאוד של מאגרי מידע וטקסט חופשי

באופן השוואתי. מורכבות היישום של המטא-ביקורת על דוחות ביקורת המדינה והאתגרים הייחודיים הכרוכים ביישומה על דוחות ביקורת המדינה בשלטון המקומי מוצגים ב שלהלן.

ההתפתחות הטכנולוגית בעשורים האחרונים אכן הביאה עימה מורכבות רבה יותר לתהליך הביקורת ולתוצריה וגם הגבירה את הקושי הכרוך בתכלול הממצאים מכמה דוחות ביקורת. אלה מעידים ביתר שאת על החיוניות שבשילוב המטא-ביקורת כפרדיגמה חדשה בביקורת המדינה, אך בד בבד מורכבות זו עלולה גם להקשות על יישומה של הפרדיגמה, בעיקר בביקורת המדינה בשלטון המקומי. לצד זאת, ההתפתחות המואצת בשנים האחרונות בתחום הבינה המלאכותית הביאה לכך שמשימות מסוימות ניתנות לביצוע ביתר יעילות ולכך שמתכננה הזמן גם לביצוען של משימות חדשות. במובן זה הבינה המלאכותית מגלמת

23 עיבוד שפה טבעית הוא תחום מחקר מדעי ויישומי הבוחן כיצד טכנולוגיה יכולה לסייע בהבנה ובעיבוד של טקסט או דיבור בשפה טבעית כדי לבצע משימות שונות.

24 מודל שפה הוא בבסיסו מודל הסתברותי המחשב את ההסתברות של מונח (token) להיות חלק "תקני" מהשפה שהוא למד. המונח הנלמד יכול לכלול מילה שלמה, חלק ממילה (כגון תחילית או סופית) או אפילו אות בודדת. המחשב לומד להשלים מונחים ממשפטים רבים, וכך לומד את מבנה השפה, הטיותה ומשמעות המילים בתוכה, והוא יכול להיות בסיס למשימות שפה רבות. ראו ענבל יחב ואביחי שריקי "אישה נעלה נעלה: מודלי עיבוד שפה טבעית בעברית" **חידושים בניהול** 127 (2023).

תרשים ג: מורכבות היישום של המטא-ביקורת על דוחות ביקורת המדינה
והאתגרים הייחודיים הנוספים ביישומה על דוחות ביקורת המדינה בשלטון המקומי



מידע מסוגים שונים, עם אפשרות לאתר מקורות מידע נוספים במרשתת²⁶. לאחר העלאתם הופכים מקורות המידע לבסיס ידע אחוד, מובנה, מקושר ובר-חיפוש המאפשר לאחזר מתוכו מידע מצטבר במהירות וביעילות. היישום מאפשר לקבוע מראש הנחיות מותאמות ל"יעד, הסגנון או התפקיד בשיחה" (להלן גם הנחיות מותאמות). סביבת העבודה ביישום היא סגורה ומאובטחת באופן יחסי (Closed Workspace), במובן זה שהמידע שמעלים אליה אינו משמש לאימון מודלים ציבוריים²⁷. כמו כן, למשתמשים שליטה מלאה על מקורות המידע המועלים

בהיקפים אדירים²⁵. מודלי השפה הגדולים הם בעלי "חלון הקשר" רחב (Large Context Window). המונח חלון הקשר מתייחס לכמות המידע שמודל השפה יכול לעבד בו-זמנית, כולל הנחיות המשתמשים, שיחות קודמות ומקורות המידע שהועלו לשיחה. ככל שחלון ההקשר רחב יותר, כך כמות המידע שהמודל יכול לעבד גדולה יותר.

NotebookLM (ובעברית: מחברת מודל שפה - Notebook Language Model) הוא יישום הנשען על יכולות מודלי שפה גדולים בעלי חלון הקשר רחב. יישום זה מציע סביבת עבודה ייעודית שאפשר לטעון אליה מאות מקורות

25 קהילת AI לומדת הטמעת בינה מלאכותית יוצרת בהוראה באקדמיה 7-8 (דו"ח מסכם של קהילת AI לומדת באוניברסיטת תל אביב 2024). אופן הפעולה של המודלים הללו מוביל לכך שהפלט שלהם עשוי להיראות משכנע, אך למעשה הוא עלול לכלול אי-דיוקים או "הזיות" (Hallucinations) של עובדות, ועשוי להיות מוטעה או אף להציג מידע שגוי לחלוטין. ראו להלן תת-הפרק "סיכונים ואתגרים ביישום המטא-ביקורת בכלי בינה מלאכותית".

26 נכון למועד כתיבת המאמר היה אפשר להעלות עד 300 מקורות מסוגים שונים: קובצי PDF או Word, סרטונים וקישורים לעמודי תוכן במרשתת.

27 בעת השימוש במודלים ציבוריים של בינה מלאכותית, ובהם מודלים שהשימוש בהם אינו כרוך בתשלום, המידע שהמשתמשים מעלים לכלי והמשוב שלהם על התשובות שקיבלו משמשים לאימון המודל ולשיפור רמת הדיוק שלו בעתיד. לכן מודלים אלה עלולים לחשוף את המשתמשים בהם לסיכונים גדולים של אבטחת מידע ופגיעה בפרטיות.

השלב השלישי התמקד בתיקוף מהימנות התשובות המתקבלות ומידת היישום של ההנחיות המותאמות. הבדיקה נעשתה באמצעות שאלות ממוקדות במתכונת של סקירה נרטיבית, קרי שאלות ממוקדות על ממצאים או על המלצות ממספר מצומצם של דוחות נבחרים שפורסמו בשנת 2025. התשובות שהתקבלו בשאלות נבדקו באופן ידני בהשוואה לדוחות הביקורת. הבדיקה והתיקוף נעשו באופן איטרטיבי, דהיינו התשובות נבדקו כנגד הדוחות, וההנחיות המותאמות עודכנו בהתאם, וחוזר חלילה, עד שהתקבלו תשובות מספקות מבחינת רמת המהימנות שלהן, הסגנון והמבנה.

בשלב הרביעי והאחרון הורחבו השאלות לחילוף תובנות-על משלושה סוגים: (א) זיהוי דפוס ליקויים חוזר או בעיות שורש בהתנהלותו של גוף מבוקר מסוים; (ב) זיהוי כשלי אסדרה או כשלים מערכתיים; (ג) זיהוי של נושאים או גופים מבוקרים שלא נכללו בביקורת כלל או שלא נכללו בביקורת זמן רב, בהתבסס על מקורות המידע שהועלו למחברת העבודה.³¹ מתוך ניתוח רפלקטיבי רב-שנתי ורב-נושאי³², ניסוח השאלות, בדומה לניסוח ההנחיות המותאמות, נעשה גם הוא בסיוע כלי בינה מלאכותית נוספים המבוססים על מודלי שפה גדולים בעלי הקשר רחב. גם תיקוף המהימנות של תובנות-העל שהתקבלו נעשה בסיוע כלי בינה מלאכותית נוספים, לרבות בדרך של "הנדסה לאחור" (Reverse Engineering)³³. שלבי העבודה מוצגים ב שלהלן.

לסביבת העבודה, באופן המפחית את הסיכון לטעויות ולשגיאות. מאפייניו של NotebookLM: סביבת עבודה סגורה ומאובטחת, חלון הקשר רחב המאפשר עיבוד של מאות מסמכים בו-זמנית ואפשרות לקבוע הנחיות מותאמות אישית לגבי היעד, התפקיד או הסגנון בשיחה, הופכים אותו למתאים לבדיקת ההיתכנות ראשונית של יישום המטא-ביקורת באמצעות כלי בינה מלאכותית.

NotebookLM כמקרה בוחן

בדיקת ההיתכנות ליישום המטא-ביקורת באמצעות NotebookLM, כמקרה בוחן, נעשתה בתהליך רב-שלבי²⁸. **בשלב הראשון** הועלו למחברת עבודה חדשה 260 דוחות ביקורת מלאים של משרד מבקר המדינה: כל הדוחות על ביקורות שנעשו בשלטון המקומי ושפורסמו באתר המרשתת של משרד מבקר המדינה בשנים 2010 - 2026²⁹. **בשלב השני** נקבעו הנחיות מותאמות במחברת העבודה בהתייחס למטרות השיחה ומבנה התשובה המחייב. ניתנה הנחיה ברורה כי כל תשובה תגובה בהפניה מדויקת לדוח הביקורת ולקטע המסוים שעליו היא מסתמכת (Grounding), כדי לאפשר למשתמשים לתקף את הממצאים באופן מיידי ובלתי אמצעי. כמו כן ניתנה הנחיה ברורה לשקף בתשובה מקרים של השמטה מכוונת של מידע מתוך הניתוח. ההנחיות המותאמות נוסחו בסיוע כלי בינה מלאכותית נוספים המבוססים על מודלי שפה גדולים בעלי הקשר רחב³⁰.

28 הבדיקה נעשתה בתקופה שבין ינואר למרץ 2026 בסביבת עבודה סגורה, בגרסת ברירת המחדל של המודל. מקורות המידע, השאלות וההנחיות המותאמות היו בעברית בלבד (שפת הקלט ושפת הפלט). המודל התבקש להשיב תשובות ארוכות יותר (Longer Response).

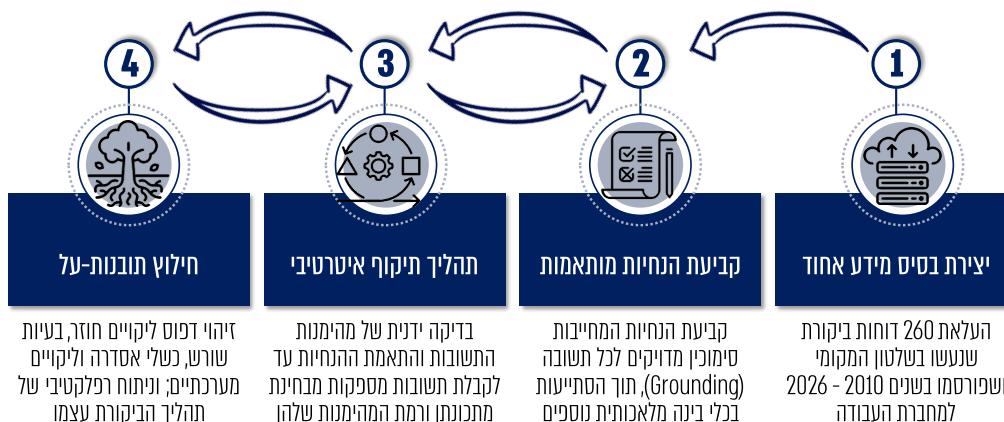
29 המידע הועלה באמצעות היפר-קישור לדוח שפורסם בפורמט PDF. כל דוח הועלה כמקור נפרד. בין המקורות נכללו דוחות רגילים, דוחות מיוחדים (לרבות דוחות ביקורת שפורסמו אחרי פרוץ מלחמת חרבות ברזל) וכן ביקורות מעקב. לא נכללו דוחות ביקורת שכללו אזכור של גופי השלטון המקומי אך לא נכתבו על ידי חטיבת הביקורת על השלטון המקומי. הדוחות הועלו בנוסח מלא, בלי להסיר פרטים מזהים של הגופים המבוקרים (Data Anonymization), שכן בדיקת ההיתכנות הראשונית התמקדה ביכולת לחלץ תובנות חדשות ברמה מספקת של מהימנות ושלמות. בחינה ותיקוף מלאים של הפרדיגמה המוצעת עשויים להצריך הסרה של פרטים המזהים של הגופים המבוקרים, כדי לזהות ולנטרל הטיות אפשריות של המודל. תקופת הזמן (2010 עד 2026) נקבעה בשל המגבלה על כמות מקורות המידע שאפשר להעלות למחברת העבודה. הועלו דוחות ביקורת שפורסמו עד סוף ינואר 2026.

30 Gemini 30 (גרסת ה-Pro) Perplexity-1 (הגרסה החינמית).

31 כדי לבחון את ההיתכנות לשימוש בבינה מלאכותית לניתוח של פערים בכיסוי גופים מבוקרים או נושאים שלא נבדקו כלל נדרש להעלות את רשימת כלל הגופים העומדים לביקורת המדינה ומסמכי יסוד על השלטון המקומי, ובכללם תוכניות עבודה של גופי האסדרה וכן מחקרים וניירות עמדה על אתגרי השלטון המקומי.

32 להלן דוגמאות כלליות לנוסחי ההנחיות שניתנו (prompts): [אתר ומפה את הדוחות על רשות מקומית X וחלץ מתוכם דפוס ליקויים חוזר ובעיות שורש]; [זהה שלוש תובנות חדשות שאינן מופיעות באף דוח בודד ושמצביעות על כשלי אסדרה או כשלים מערכתיים]; [נתח דפוסים בבחירת נושאי הביקורת בשנים 2010 - 2026]. יש לציין כי הוחלט שלא לכלול בבדיקה את השימוש במטא-ביקורת לתכנון ביקורות חדשות, משום שניתוח כזה הוא פשוט יותר, והמחשת ההיתכנות הראשונית אינה מחייבת בדיקות מעמיקות ומורכבות.

33 המודלים התבקשו לבדוק אם אפשר לבסס את התובנה שהתקבלה על המקורות הקיימים.



כמה פעמים במתכונת דומה, ותוצאותיהן נבדקו ידנית לתיקוף של מהימנותן ושלמותן. תוצאות הבדיקה מפורטות להלן.

דפוס ליקויים חוזר ובעיות שורש בהתנהלות רשות מקומית מסוימת

בפעם הראשונה ניתנה הנחיה אחת דו-שלבית: לאתר ולמפות את דוחות הביקורת שבהם מוזכרת רשות מקומית מסוימת ולנתח אותם, לצורך זיהוי דפוס ליקויים חוזר ובעיות שורש באותה הרשות. בפעם השנייה ההנחיה ניתנה בהדרגה: תחילה לאתר את דוחות הביקורת, ורק לאחר השלמת המיפוי הראשוני ומתן הנחיית המשך, לעשות ניתוח מתכלל של הממצאים מתוך הדוחות שאותרו. בבדיקה הידנית נמצא ביסוס מספק לממצאי הניתוח המתכלל ולתובנות שהוא העלה, אך נמצא גם כי מהמיפוי הראשוני הושמטו בשוגג דוחות מסוימים (False Negative)³⁴.

כדי להגדיל את עצמאות הממצאים וכדי לצמצם את ההשפעה של הנחיות קודמות שניתנו למודל בתוך השיחה על תוצרי הניתוח (History Echoes)³⁴, נמחקה היסטוריית השיחה באופן מלא בכל פעם מחדש לפני שניתנה הנחיה חדשה. כמו כן המודל התבקש לחוות את דעתו בנוגע למידת הביסוס של תוצרי הניתוח שלו עצמו במקורות המידע שהועלו (Self-Consistency), גם כן לאחר מחיקת היסטוריית השיחה. בהינתן פעולות אלה, העובדה שהניתוח המתכלל העלה תובנות דומות בכל פעם, כפי שיפורט בהרחבה להלן, מחזקת במידה מסוימת את מהימנות הממצאים.

סקירת תוצאות בדיקת ההיתכנות הראשונית

הבדיקה נעשתה בנפרד לכל אחת מתובנות-העל, והיא נעשתה בשני אופנים: כהנחיה אחת רב-שלבית וכהנחיה מדורגת (Chain of Thought Prompting)³⁵. ההנחיות ניתנו

34 ראו Adi Simhi et al., Old Habits Die Hard: How Conversational History Geometrically Traps LLMs, Arxiv (February 2026) ARXIV:2603.03308v1. [cs.CL].

35 שיטה זו נועדה לאפשר למודלי שפה גדולים להתמודד עם משימות מורכבות או ארוכות או כאלו שדורשות דיוק רב, כלומר משימות המייצרות עומס קוגניטיבי על המודל באמצעות פירוקן ליחידות מידע קטנות וממוקדות. ראו in Jason Wei et al, Chain-of-Thought Prompting Elicits Reasoning in Large Language Models, ARXIV (January 2022) ARXIV:2201.11903 [cs.CL].

36 תוצאות המיפוי הראשוני כללו במצטבר כ-20 דוחות ביקורת (בהתייחס לכלל התשובות שהתקבלו). בדיקה ידנית שנעשתה בהתייחס לדוחות שפורסמו בשנים 2018 - 2026 העלתה כי הושמטו שני דוחות ביקורת שפורסמו בשנים הללו וכללו ממצאים בנוגע לרשות המקומית המסוימת (משישה דוחות, כלומר כשליש מהדוחות). המודל השיב שאחד מהדוחות הושמט בשוגג, ובהתייחס לדוח האחר הוא השיב שהדוח לא נכלל מלכתחילה בין מקורות המידע שהועלו למחברת העבודה, אף על פי שטענה זו אינה נכונה.

כשלי אסדרה וכשלים מערכתיים

בתחילה ניתנה הנחיה אחת דו-שלבית ולאחריה הנחיה מדורגת: לאתר ולמפות את דוחות הביקורת הכוללים ממצאים על כשלי אסדרה או כשלים מערכתיים, ומהם לזהות שלוש תובנות-על חדשות. התוצאות נבדקו ידנית לצורך תיקוף מהימנותן. גם הפעם ממצאי הניתוח והתובנות שהועלו היו דומים בין הפעם הראשונה, הפעם השנייה ויתר הפעמים שבהם ניתנה ההנחיה. כך גם נמצא להם ביסוס מספק, אף על פי שדוחות הביקורת שעליהם התבססו היו בתחומי אסדרה שונים בכל פעם. נוסף על כך, מורכבות התובנות - שהיו דומות במהותן אך ניסוחן השתנה מפעם לפעם - חייבה בקרה ידנית קפדנית והפעלת שיקול דעת מקצועי כדי שיהיה אפשר לחלץ מהן רעיון סדור אחד.

להלן יוצגו דוגמאות נבחרות של כשלי אסדרה וכשלים מערכתיים שהועלו בניתוח:

אשליית הקדמה הטכנולוגית: ניתוח מתכלל של ממצאים מתוך שלושה דוחות ביקורת, כל אחד בתחום אסדרה שונה³⁷, העלה כי הרשויות המקומיות הטמיעו בעבודתן מערכות טכנולוגיות מתקדמות המסוגלות לעבד ולנתח נתונים בהיקף נרחב ולשימושים מגוונים. עם זאת, לא מנוהלים במערכות הללו נתונים ומידע בסיסיים החיוניים לתהליכי קבלת ההחלטות ולניהול סיכונים. משכך, נתונים אלה גם אינם משמשים הלכה למעשה בתהליכי קבלת ההחלטות.

הגם שנמצאה שונות ברשימת הדוחות שנכללו בניתוח ובדוחות שהושמטו בשוגג מהמיכוי הראשוני, היו הממצאים והתובנות דומים בעיקרם בפעם הראשונה ובפעם השנייה וביתר הפעמים שבהן ניתנה ההנחיה, והם התמקדו בהשוואה בין שלוש קבוצות של נושאים. הניתוח הציע לקשור בין ממצאי ביקורת מוקדמים על ביטוח נכסי ציבור, מבנים שאינם בשימוש ותחום התכנון והבנייה וממצאים מאוחרים על ניהול נכסי מקרקעין (מהשנים 2012, 2016, 2019 ו-2023)³⁷; בין ממצאים מוקדמים על החוסן הפיננסי וממצאים מאוחרים על ההתנהלות בתחום היטלי ההשבחה (מהשנים 2019 ו-2024); ובין ממצאים מוקדמים וממצאים מאוחרים בתחום המכרזים וההתקשרויות (מהשנים 2014, 2015 ו-2019). לפי הניתוח, הממצאים המאוחרים בכל אחת מקבוצות הנושאים מצביעים על ליקויים בניהול הידע על המרחב הפיזי והתכנוני של הרשות המקומית המסוימת שנבדקה, על ליקויים בהקפדה על כללי המינהל התקין בתחום הרכש וההתקשרויות ועל חולשה ניהולית בתחום הפיננסי³⁸. לפי הניתוח המתכלל, גם בלי שנעשתה ביקורת מעקב שיטתית בנושאים הללו, וגם אם ליקויים פרטניים שנמצאו בעבר תוקנו לפחות בחלקם, קיים יסוד להניח כי לא חל שינוי בתרבות הארגונית של הרשות המקומית המסוימת שנבדקה בנושאים האמורים.

37 ממצאי ביקורת מוקדמים הצביעו על כך שהרשות המקומית אינה מחזיקה בנתונים על שטחם המבונה של כלל נכסי המקרקעין שלה, כדי לשערך אותם ולקבוע את היקף הכיסוי הביטוחי הדרוש לה, וכן על כך שלא ניהלה את ספר הנכסים כנדרש בדיון ולא הסדירה את ניהול הנכסים ורישומם. ממצאים מאוחרים הצביעו על כך שנתו פערם באופן ניהול הנתונים על נכסי המקרקעין בתחומה של הרשות המקומית. בבדיקה בשיטת ההנדסה לאחר המודל הפנה לדוח נוסף בנושא הצללת המרחב הציבורי כחיוזוק נוסף לממצא המתכלל.

38 ממצאי הביקורת הצביעו על היעדר מיצוי של תקציב שהועמד לרשות הרשות המקומית או על אובדן הכנסות פוטנציאליות. בניתוח הוזכרו שלושה דוחות נוספים, נקודתיים ומערכתיים, שהצביעו על ליקויים גם הם, בבקרה התקציבית של הרשות המקומית או התאגידים העירוניים המשוכים אליה בהתקשרויות עם נתוני שירותים חיצוניים.

39 בשאלתה הראשונה נכללו הדוחות האלה: מבקר המדינה **ניהול מערכות מידע ברשויות המקומיות** (2022); מבקר המדינה **ניהול של נכסי מקרקעין בידי רשויות מקומיות** (2023); מבקר המדינה **פעילות השירות הפסיכולוגי-חינוכי למענה רגשי ונפשי בשגרה ובמלחמת חרבות ברזל** (2025). בשאלתה השנייה נכללו דוחות בתחומי אסדרה נוספים.

ריק רגולטורי: ניתוח מתכלל של ממצאים מתוך שלושה דוחות ביקורת המתייחסים למצב של היעדר אסדרה או קיומה של אסדרה חלקית או עמומה (Fuzzy Control)⁴⁰ בתחומי הביקורת השונים העלה כי מצב זה חוצה תחומי אסדרה וחוצה גם גופים מבוקרים⁴¹. הוא מייצר "ריק רגולטורי" העלול להוביל רשויות מקומיות לגבש מסמכי מדיניות או לקבוע נוהלי עבודה פנימיים ללא תוקף משפטי מחייב. בעקבות זאת עלולה התנהלות להתאפיין בחוסר אחידות, וגובר החשש לפגיעה בעקרונות השוויון והמינהל התקין, וכך גם החשיפה למעילות ולהונאות. גם פעולות האכיפה מצד גופי האסדרה עלולות להתאפיין בחוסר אחידות בהיעדרה של מסגרת מאסדרת מחייבת⁴².

איים של מידע: ניתוח מתכלל של ממצאים מתוך שלושה דוחות ביקורת בתחומי החינוך, הרווחה והשירותים החברתיים וכן בתחומי הבריאות והסביבה⁴³ העלה כי העברת האחריות על אספקת שירותים בתחומים אלה לנתני שירותים חיצוניים, לתאגידים עירוניים או לאיגודי ערים עלולה להוביל להיווצרות "איים של מידע" (Data Silos), מצב שבו המידע על צורכיהם של תושבי הרשויות המקומיות והשירותים הניתנים להם מקוטע ומבוזר. מצב זה עלול להוביל לאובדן הגישה של הרשויות המקומיות למידע חיוני זה, ובעקבות זאת ליטול מהן את היכולת לקבל תמונה שלמה ומלאה על צורכי התושבים ועל השירותים הניתנים

להם ולקבל החלטות מושכלות ומבוססות נתונים בעיניים. תוצרי הניתוח משלבים סקירה נרטיבית של ממצאים קיימים לצד הרחבה שלהם, ולא כולם בהכרח מציגים נקודת מבט חדשה לחלוטין. אולם התוצרים כללו גם תובנה מצרפית על מצבים העלולים להשפיע על יציבותן הארגונית של הרשויות המקומיות בעיתות משבר וחירום ולכנוע ברציפות תפקודן: הישענות על מערכות מידע שאינן מגובות בנתונים ומידע החיוניים לקבלת ההחלטות; מיקור חוץ של מנגנוני הבקרה והפיקוח על נותני שירותים חיצוניים ועל התאגידים העירוניים ואיגודי הערים; וכן ריק רגולטורי בתחומי פעילות נרחבים. זו אינה רק הרחבה של ממצאים קיימים; מדובר בחיבור בין ממצאים מתוך דוחות נפרדים לכדי תובנה קוהרנטית אחת.

כאשר ניתנה למודל הנחיה מכורשת לכלול בניתוח מנגנון בקרה עצמית על רמת החידוש שהוא מציע, לצד שילובן של טכניקות נוספות לשיפור המהימנות והשלמות של התוצאות, הוא הציע תובנה מצרפית נוספת. תובנה זו מציעה הרחבה של תובנות אחרות בעניין יכולת ניהול הנתונים של הרשויות המקומיות והיעדרה של אסדרה ממצה בתחומים שונים. ניתוח מתכלל של ממצאים מתוך עשרה דוחות ביקורת בתחומי אסדרה מגוונים⁴⁴ הצביע על קיומו של **כשל רגולטורי מעגלי**: המדינה מבזרת סמכויות לשלטון המקומי לעיתים ללא מסגרת אסדרה ממצה ובלי להסדיר את יכולת השלטון המקומי לנהל את המידע

40 המונח Fuzzy Control מתייחס להתפתחות האסדרה ביחסי השלטון המרכזי והמקומי בישראל, ולכן שהנורמות המטושטשות ביחסים אלה הובילו להתפתחות של השלטון המקומי ללא יד מכוונת. ראו David Dery, Fuzzy control, Journal of Public Administration Research and Theory 12 191 (2002).

41 בשאלתה הראשונה נכללו דוחות בתחומים של תכנון ובנייה, רכש והתקשרויות וניהול פיננסי. בשאלתה השנייה נכללו דוחות נוספים באותם תחומי אסדרה.

42 באחת מהשאלות התשובה כללה התייחסות גם לתופעה של "רגולציה גנרית" (Flat Regulation), מצב שבו מנגנוני האסדרה אינם מייצרים שונות (דיפרנציאציה) בין גופים שונים (לדוגמה על בסיס מצב פיננסי או מאפייני האוכלוסייה), ולכן הם מייצרים ומשמרים אי-שוויון. תופעה זו חוצה נושאים, ואפשר לזהותה בתחומי השירותים החברתיים והרווחה, ראו מבקר המדינה **מימון שירותי רווחה** (2021) וכן בתחום הניהול הפיננסי, ראו מבקר המדינה **תוכניות הבראה ברשויות המקומיות** (2023).

43 מבקר המדינה **אסדרת אספקתם של שירותי הרווחה בקהילה על ידי הרשויות המקומיות** (2022); מבקר המדינה **פעילות השירות הפסיכולוגי-חינוכי למענה רגשי ונפשי בשגרה ובמלחמת חרבות ברזל** (2025); מבקר המדינה **הפיקוח על בשר ומוצרי - השירות הווטרינרי ברשויות המקומיות** (2023). בשאלתה השנייה נכללו דוחות בתחומי אסדרה נוספים.

44 המודל התבסס לזהות דוחות ביקורת שבהם מוזכרים כשלי אסדרה, ובהתבסס על הממצאים שבהם נלסח תובנות-על חדשות. להלן פירוט הדוחות שהמודל איתר וכלל בניתוח: בנושא רווחה ושירותים חברתיים - מבקר המדינה **טיפול הרשויות בתופעה של ניצול קטנים לזנות** (2014) ומבקר המדינה **ההתמודדות עם תופעת האלימות בין בני זוג** (2021); בנושא תכנון ובנייה - מבקר המדינה **עבירות בנייה של רשויות מקומיות ותאגידים עירוניים** (2017) ומבקר המדינה **ייעול הליכי רישוי הבנייה** (2021); בנושא סדירות, חוקיות ומינהל תקין - מבקר המדינה **ניהול הרשומות האלקטרוניות ברשויות המקומיות** (2014), **דירקטוריונים של חברות עירוניות** (2017) ומבקר המדינה **המבקר ברשות המקומית - תפקידו והתנהלותו** (2023); בנושא מערכות מידע וסייבר - מבקר המדינה **אבטחת מידע והגנת הפרטיות** (2017); בנושא חירום וביטחון - מבקר המדינה **התמודדות הרשויות המקומיות עם משבר הקורונה** (2021) ומבקר המדינה **טיפול הרשויות המקומיות במבנים מסוכנים - ביקורת מעקב** (2025).

תוצאות ההנחיה המדורגת - שלפיהן הנושאים הנבדקים בתדירות הגבוהה ביותר הם תכנון ורישוי בנייה, ניהול כספי ורווחה ושירותים חברתיים - עולות בקנה אחד עם בדיקה ידנית של התפלגות נושאי הביקורת בשנים 2018 - 2026. אף על פי שתוצאות הניתוח כללו כימות משוער של הדוחות בנושאים השונים, פירוט של השנים שבהן הם פורסמו ודוגמאות לדוחות שנכללו בסיווגים השונים, לא כלל הפלט סיווג מלא של הדוחות, ולכן לא התאפשרה בקרה מלאה על תוצאות הניתוח. זוהי המחשה של היעדר השקיפות ובעיית ההסבריות (Explainability) המיוחסים לשימוש בבינה מלאכותית⁴⁶.

דפוסים נוספים בבחירת נושאי הביקורת: הניתוח הצביע על כך שנושאים של חירום וביטחון נבדקים באופן תגובתי ומוטה משברים, ואילו בעיתות של שגרה נושאים אלה נבדקים בתדירות נמוכה. הם חוזרים למוקד הביקורת רק בעקבות משברים לאומיים. ניתוח נפרד, שהתמקד בשנים 2024 - 2025, העלה כי חל שינוי מסוים בדפוס זה על רקע מלחמת חרבות ברזל: חלק ניכר מדוחות הביקורת החלו לשלב ממד של חירום וביטחון, כחלק פלוג בביקורת ובלתי נפרד ממנה, ולא כנושא לביקורת כשלעצמו. הניתוח הצביע גם על כך שישנם תחומים שנמצאו בהם כשלים משמעותיים או כשלי אסדרה בנושא מסוים, אך רק כעבור פרק זמן ממושך מאוד הנושא נבדק בשנית, בפני עצמו או כחלק מנושא ביקורת אחר⁴⁷.

דפוסים בבחירת הגופים המבוקרים: הניתוח הצביע על הטיה מסוימת במאפייניהן של הרשויות המקומיות הנכללות בביקורת, בשים לב לגודלן, מצבן הפיננסי ומיקומן הגיאוגרפי. לפי ממצאי הניתוח, ישנה נטייה לכלול בביקורת העומק רשויות מקומיות המצויות בקצוות: מחד גיסא רשויות גדולות המשתייכות למדד חברתי-כלכלי גבוה, ולכן היקף התקציבים שהן מנהלות גדול יותר וההשלכות של ליקויים בהתנהלותן נרחבים יותר, ומאידך גיסא רשויות המשתייכות למדד חברתי-כלכלי נמוך, הממוקמות בפריפריה ושלרוב תושביהן משתייכים למגזר הלא-יהודי. רשויות אלה לוקות בכשלים

והנתונים בנושאים שהאצילה⁴⁵. כך נוצר מעגל קסמים שבו לשלטון המרכזי אין גישה לנתונים ההכרחיים לקביעת מסגרת אסדרה ממצה להתנהלות השלטון המקומי ותקצוב מותאם, וחוזר חלילה.

ניתוח רפלקטיבי של תהליך הביקורת עצמו

בפעם הראשונה ניתנה הנחיה אחת שבה התבקש ניתוח מתכלל שמטרתו לזהות דפוסים והטיות בבחירת נושאי הביקורת ובבחירת הגופים המבוקרים. מבין הסוגים השונים של תובנות-העל, התובנות מסוג זה מצריכות את הבקרה האנושית ההדוקה והקפדנית ביותר, ויישומה הוא גם המורכב ביותר. מורכבות זו עשויה להסביר את השונות הרבה בין תוצאות השאלית הראשונה (הרב-שלבית) והשנייה (המדורגת). להלן יפורטו כמה מבין תובנות-העל שהניתוח העלה.

דפוסים בבחירת נושאי הביקורת: כאשר ניתנה הנחיה אחת רב-שלבית הניתוח הצביע על שלושה נושאי ליבה הנבדקים בתדירות הגבוהה ביותר: תכנון ורישוי בנייה, ניהול כספי ורכש והתקשרויות. נוסף על כך הניתוח הצביע על עיסוק גובר בשנים האחרונות בתחום הדיגיטציה ומערכות המידע, בדגש על אבטחת מידע ושירותים מקוונים. כאשר ניתנה הנחיה מדורגת, עלה מתוצאות הניתוח שבשל עיסוק גובר בשנים האחרונות בנושא הרווחה והשירותים החברתיים, הוא נבדק בתדירות גבוהה יותר מנושא הרכש וההתקשרויות. אולם גם תוצאות ההנחיה המדורגת הצביעו על מיקוד מערכתי קבוע ועקבי בשני הנושאים האחרים: תכנון ורישוי בנייה וניהול כספי. כאשר ניתנה הנחיה מפורשת לנתח את הדפוסים בבחירת נושאי הביקורת, העלה הניתוח אפשרות כי בחירת הנושאים מושפעת מהתמורות שחלו בתחום ביקורת המדינה, ובעיקר מהמעבר מביקורת על החוקיות והסדירות של הניהול הכספי של הגופים המבוקרים (ביקורת ציות) לביקורת על היעילות, המועילות והיחסות בהתנהלותם (ביקורת ביצועים).

45 הכוונה היא, לדוגמה, למצב שבו הרשויות המקומיות מספקות שירותים לתושב, כזרועה הארוכה של המדינה, בלי שהיא מנחה אותן לנהל את הנתונים על השירותים שהן מספקות או לדווח לה עליהם. כפועל יוצא, המדינה אינה יכולה להשתמש בנתונים להערכת המדיניות ולתיקופה או לבחינה מחדש ולעדכון הנחיותיה לרשויות המקומיות בנושא. לעניין החשיבות של הערכת מדיניות ציבורית, בעת מעשה ולאחר מעשה, ובפרט בעיתות משבר וחירום ראו רון צור וניסים (נס) כהן **הערכת מדיניות בשעת משבר: מודל להפקת לקחים בעקבות משבר הקורונה** (2021) וכן רון צור וניסים כהן, "אי-משילות בשעת משבר: העדר הערכת מדיניות והשפעתה על התמודדותה של ישראל עם הקורונה - מאמר דעה", **ביטחון סוציאלי** 115 11 (2021).

46 בעיית ההסבריות מתייחסת לקושי לחזות את פעולותיה של הבינה המלאכותית ולמורכבות הכרוכה בהסברת הלוגיקה של קבלת ההחלטות שלה לאחר מעשה. ראו Brent D. Mittelstadt et al., *The Ethics of Algorithms: Mapping the Debate*, 1 *Big Data & Society* 3 (2016).

47 כך למשל, נושא הביקורת החיצונית ברשויות המקומיות באמצעות רואי חשבון חיצוניים נבדק לראשונה (בטווח השנים שנכללו בניתוח) בשנת 2015. בשנת 2023 הנושא נכלל בביקורת בנושא אחר, והועלה כי לא תוקנו ליקויים שהועלו עוד בשנת 2015; נושא הממשל התאגדי ודירקטוריונים בתאגידים עירוניים נבדק לראשונה (בטווח השנים שנכללו בניתוח) בשנת 2017. בשנת 2025 הנושא נכלל בביקורת בנושא אחר, והועלה כי לא תוקנו ליקויים שהועלו עוד בשנת 2017.

ניהוליים, באי-סדרים ובהיעדר תשתיות בסיסיות. לפי הניתוח, רשויות מקומיות שאינן בקצוות הללו נוטות להיכלל פחות בביקורות העומק, אך נכללות בהשוואות שנעשות ברמה הכלל-ארצית⁴⁸.

יש לציין שבבחירת הרשויות המקומיות לביקורת, במיוחד בשנים האחרונות, מקפידים ככל האפשר על תמהיל של כמה מאפיינים לצד גודלן ושיכון למדד החברתי-כלכלי⁴⁹. לפי המודל, ייתכן שההסבר לממצאי הניתוח קשור בכך שרשויות מקומיות בקצוות מקבלות לרוב בולטות רבה יותר בדוח - לטובה או לרעה - והמודל מפרש זאת כהטיה מסוימת בהליך הבחירה⁵⁰. מכל מקום, ממצאים אלה מחייבים המשך בדיקה והעמקה שלא התאפשרו במסגרת בדיקת ההיתכנות הראשונית⁵¹.

ניתוח תוצאות בדיקת ההיתכנות הראשונית

לצורך הערכת תרומתן של תובנות-העל שהתקבלו נקבע מדרג לפי רמת החידוש המתבטאת בהן: תובנה המזהה דפוסים בממצאים, בנושאים או בגופים המבוקרים (להלן גם תובנה נרטיבית מצומצמת) סווגה ברמת החידוש הנמוכה ביותר. תובנה המתכללת ממצאים ממקבץ דוחות ומחלצת מהם דפוס ליקויים מערכתי (להלן גם תובנה נרטיבית מורחבת) סווגה ברמת חידוש בינונית. תובנה המזהה מתאם בין נושאים שלא בא לידי ביטוי באף דוח בודד ומגלמת מסקנה אחת קוהרנטית (להלן גם תובנה מצרפית) סווגה ברמת החידוש הגבוהה ביותר.

תובנות-העל שהתקבלו סווגו גם לפי רמת המורכבות של הבקרה האנושית הנדרשת לתיקוף של מהימנותן ושלמותן: ככל שהבקרה הייתה כרוכה בבדיקת כמות גדולה יותר של דוחות ביקורת (בחינת השמטה של דוחות וביסוס הממצאים מתוכם), כך היא הוגדרה כמורכבת יותר; בקרה הכרוכה בבדיקה של דוחות בודדים סווגה ברמת המורכבות הנמוכה ביותר; בקרה הכרוכה בבדיקה של עשרות בודדות של דוחות סווגה ברמת מורכבות בינונית;

ואילו בקרה הכרוכה בבדיקה של כלל הדוחות או של עשרות רבות מהם סווגה ברמת המורכבות הגבוהה ביותר. תובנות-העל שעלו בבדיקת ההיתכנות, לפי סוגיהן, רמת החידוש בהן ומורכבות הבקרה האנושית עליהן לתיקוף של מהימנותן ושלמותן, מוצגות ב 2 שלהלן.

בדיקת ההיתכנות הראשונית מצביעה על כך ששילוב כלי בינה מלאכותית בתהליך המטא-ביקורת מאפשר לא רק לייצר סקירה נרטיבית של ממצאים, אלא גם להרחיב אותם. תהליך זה מסוגל אף לייצר תובנות-על מצרפיות, החוצות תחומי אסדרה וגופים מבוקרים, תובנות שלא היה אפשר לזהות בקריאה של דוחות בודדים. את התובנות הללו אפשר להטמיע בתהליכי העבודה הפנימיים של משרד מבקר המדינה: בקביעת תוכנית העבודה השנתית, בתהליכי הפקת לקחים ובהחלטה על נושאים שבהם יתקיימו ביקורות מעקב. אפשר גם לפרסם את התובנות שהתקבלו באופן עיתי לשימוש קובעי המדיניות, מקבלי ההחלטות והציבור בכללותו. הטכניקות השונות ששולבו בבדיקה ממחישות את היותה של הבינה המלאכותית מכפיל כוח - לולא ההכוונה האנושית והתיקוף האנושי האיטרטיביים (חוזרים ונשנים), כפי שתוארו לעיל, התוצרים שהיו מתקבלים מהניתוח המתכלל היו מהימנים פחות, שלמים פחות וחשוב מכול מהדשים פחות.

האמור כאן מחדד את האתגרים ואת הסיכונים המובנים ביישום הפרדיגמה המוצעת בכלי בינה מלאכותית. הפרק הבא במאמר יבקש להתמודד עם הסיכונים הללו ולהציע מסגרת עבודה שתאפשר שילוב אחראי וזהיר של כלי בינה מלאכותית בתהליך המטא-ביקורת.

סיכונים ואתגרים ביישום המטא-ביקורת בכלי בינה מלאכותית

שילוב כלי בינה מלאכותית בעבודת הביקורת, ובפרט כחלק מיישום המטא-ביקורת, כרוך בסיכונים מובנים

48 ביקורות מערכתיות על גופי השלטון המקומי כוללות לרוב שלושה רובדי בדיקה: ברובד הבדיקה העליון נעשים עיבודים וניתוחים של נתונים על כלל הרשויות המקומיות שזמינותם גבוהה. עיבוד נתונים אלה וניתוחם מאפשרים בחינה רחבה וכלל-ארצית של נושא הביקורת הנבדק, תוך התייחסות למאפיינים ייחודיים שעשויים להגביר את חשיפתן של רשויות מסוימות לסיכונים בנושא זה. ברובד הבדיקה האמצעי נכללת קבוצה מצומצמת יותר של רשויות מקומיות, ונבדקים נתונים שאינם פומביים או שזמינותם אינה גבוהה. ברובד הבדיקה התחתון נעשים עיבודים וניתוחים של נתונים מפורטים מאוד שאינם פומביים ולכן זמינותם אינה גבוהה - זוהי ביקורת העומק. ראו רייכמן-עמית "ללמוד לבחור נכון", לעיל ה'ש 19.

49 לעיתים קרובות מצוין בדוח כי הרשויות המקומיות שבהן התבצעה הביקורת נבחרו על פי תמהיל של כמה מאפיינים, ובהם המעמד המוניציפלי של הרשות (עירייה, מועצה מקומית או מועצה אזורית), האזור הגיאוגרפי שבו שוכנת הרשות (דרום, מרכז או צפון), מספר תושביה והמגזר שאליו משתייכים מרביתם. לעיתים משתמשים בהליך הבחירה בנתונים נוספים הקשורים במישרין לנושא הביקורת.

50 הסבר נוסף עשוי להיות קשור במגבלות המובנות של תהליך הבחירה של הרשויות המקומיות לביקורת, שכן כאשר הביקורת נעשית במספר מצומצם של רשויות מקומיות הרצון לייצר תמהיל של רשויות מקומיות עלול להוביל לבחירה של רשויות בקצוות.

51 בדיקה כזו מחייבת מיפוי וניתוח של מאפייניהם של הגופים המבוקרים שנכללו בביקורת ורמת ההעמקה של הביקורת שנעשתה בהם, ללא הסתייעות בכלי בינה מלאכותית.

לוח 2: תובנות-העל שעלו בבדיקת ההיתכנות הראשונית, לפי רמת החידוש בהן ומורכבות הבקרה האנושית עליהן

תובנת-העל	פירוט תובנת-העל שהתקבלה	סיווג התובנה	רמת החידוש	מורכבות הבקרה
דפוס ליקויים חוזר ובעיות שורש	זיהוי מתאם בין ממצאים מתחומי ביקורת שונים בנוגע לגוף מבוקר מסוים: איתור ממצאים בנושאים דומים מתקופות זמן שונות (לדוגמה, ביטוח נכסים, טיפול במבנים מסוכנים וניהול נכסי מקרקעין; חוסן פיננסי והיטלי השבחה)	נרטיבית מצומצמת	●	●
כשלי אסדרה וליקויים מערכתיים	זיהוי דפוס מערכתי בממצאים מתחומי ביקורת שונים: הכללה של הממצאים המתכללים לכדי דפוס רחב, מעבר לדוח הבודד (לדוגמה, ריק רגולטורי; אשליית הקדמה הטכנולוגית; איים של מידע בשל מיקור חוץ נרחב)	נרטיבית מורחבת	●	●
זיהוי מתאם רב-תחומי בין דפוסים מערכתיים שלא מופיע באף דוח בודד והפיכתו לכדי תובנה קוהרנטית אחת (לדוגמה, הקשר בין ריק רגולטורי, אשליית הקדמה הטכנולוגית ומיקור חוץ לבין התמודדות הרשויות המקומיות בעיתות משבר וחירום; כשל רגולטורי מעגלי)		מצרפית	●	●
ניתוח רפלקטיבי של תהליך הביקורת	דפוסים בבחירת נושאי הביקורת או הגופים המבוקרים: ספירה כמותית של דוחות, לפי נושא או לפי אזכור הגופים המבוקרים	נרטיבית מצומצמת	●	●

ניתוח של דוחות ביקורת שפורסמו במרשתת, אינו כרוך בסיכון רב, ומכל מקום דרכי ההתמודדות עימו אינן ייחודיות לתהליך זה⁵³. לצד זאת ישנם סיכונים משמעותיים אחרים, המחייבים בקרה אנושית קפדנית והדוקה.

הקשורים בתהליך השליפה והניתוח של המידע ובתיקוף של מהימנות התוצאות ושלמותן, וכן בשמירה על הפרטיות ואבטחת המידע⁵². בכל הקשור לשמירת הפרטיות ואבטחת המידע תהליך העבודה המוצע, המתבסס על

52 בהקשר זה ראו Julia Kokina et al., Challenges and Opportunities for Artificial Intelligence in Auditing: Evidence from the Field, 56 International Journal of Accounting Information Systems (2025) הסיכונים העיקריים בהקשר של יישום מטא-ביקורת בכלי בינה מלאכותית הם סיכונים סיסטמיים שהתממשותם במצטבר ולאורך זמן עלולה להסב נזק חברתי, ולא סיכונים אינדיבידואליים שהתממשותם עלולה להסב נזק ישיר לאדם מסוים. לעניין הבחנה זו ראו מיכל שור-עופרי "מעבר לשקיפות: ריבוי ומגוון במדיניות בינה מלאכותית" משפט חברה ותרבות ח: בינה מלאכותית במשפט הישראלי 179 (2025).

53 קיים סיכון מסוים של חשיפת מידע פנים-ארגוני על שיטות עבודה, אם המטא-ביקורת אכן תשמש בעתיד לקביעת תוכנית העבודה השנתית או תוכנית העבודה לביקורת בנושא מסוים. אך משום שמידע זה אינו רגיש, והוא מבוסס על מידע שפורסם במרשתת, הנזק העלול להיגרם מחשיפתו אינו רב. לכן יש להוסיף כי אפשר לשתף את מחברת העבודה עם משתמשים נוספים, אך היסטוריית החיפוש היא פרטית.

את הסיכונים והאתגרים העיקריים הקשורים ביישום המטא-ביקורת בכלי בינה מלאכותית אפשר לחלק לשלוש קטגוריות השלובות זו בזו: (א) סיכונים בשלב שלילת המידע, ובפרט השמטה של מידע בשוגג (False Negative); (ב) סיכונים בשלב ניתוח המידע, ובכללם תופעת ה"אפופניה" (Apophenia)⁵⁴ והחשש מהזיות (AI Hallucinations)⁵⁵ והטיות אלגוריתמיות (Algorithmic Bias)⁵⁶; (ג) אתגרים בשלב התיקוף של מהימנות התוצאות ושלמותן, ובכללם בעיית ההסבריות (Explainability) ושאלת האחריות. להלן יפורטו בהרחבה הסיכונים והאתגרים הללו.

סיכונים בשליפת המידע (Retrieval) ובניתוח שלו (Processing)

שלב השליפה מתייחס לעיתוי שבו ממופים ומאותרים מקורות המידע שישמשו את המודל בהמשך, בשלב ניתוח המידע ובהפקת תובנות. הסיכונים בשלבים אלה עשויים לנבוע משתי סיבות עיקריות:

מגבלות וסיכונים המובנים במודל: מודלי שפה גדולים בעלי חלון הקשר רחב נוטים להעניק משקל עודף למידע המצוי בקצוות הקלט על חשבון מידע המצוי במרכז (Lost in the Middle). מכיוון שמודלים אלה מבוססים על חיזוי סטטיסטי, ככל שקלט המידע ארוך יותר ונדרש לאתר ולמפות מידע

מדויק מתוכו, כך גם פוחתת רמת הדיוק בתהליך השליפה (ובהמשך גם בתהליך הניתוח)⁵⁷. מגבלה נוספת קשורה במגבלות הגודל של חלון ההקשר ובמגבלות מובנות של אורך הקלט והפלט⁵⁸ המביאות לכך שהתשובה שהמודל נותן אינה בהכרח מלאה.

ארבעה סיכונים מובנים נוספים הקשורים במאפייניהם של מודלי שפה גדולים בעלי חלון הקשר רחב קיימים בשלב הניתוח של המידע: (א) תופעת ה"אפופניה" (Apophenia), המתעצמת ככל שקלט המידע גדול יותר, ושבה המודל מזהה דפוסים ומתאם (Correlation) שגויים⁵⁹; (ב) תופעת ההזיות (Hallucinations), שמביאה להמצאה של מידע שאינו מעוגן במקורות; (ג) הטיות אלגוריתמיות (Algorithmic Bias) ויצירה של "תיבות תהודה" (Echo-Chambers) המהדהדות את אותן הטיות, ובכך עלולות להביא לשעתוק של דפוסי עבר⁶⁰; (ד) נוסף על כל אלה, מודלי שפה גדולים נוטים להעדיף אחידות ותבניות על פני ריבוי וגיוון, ובעקבות זאת מציגים תמונת עולם ממורכזת וצרה מכפי שהיא בפועל ומטשטשים את קיומן של אפשרויות תשובה נוספות⁶¹.

אתגרים הקשורים באיכות קלט המידע ומורכבותו: דוחות הביקורת בנויים מטקסט בלתי מובנה (Unstructured Data)

54 המונח "אפופניה" מתאר מצב שבו מזוהים דפוסים או קשרים בתוך מידע אקראי או חסר-משמעות, בלי שדפוסים או קשרים אלה אכן קיימים. את המונח טבע לראשונה בשנת 1958 הניירולוג והפסיכיאטר קלאוס קונרד (Conrad). ראו "apophenia" Rogers Kara, Encyclopedia Britannica, www.britannica.com/topic/apophenia.

55 המונח הזיה, בהקשר לשימוש בכלי בינה מלאכותית, מתאר, בהשאלה מן הפסיכולוגיה, תופעה שבה מערכת בינה מלאכותית יוצרת תוכן המכיל מידע כוזב או מטעה המוצג כעובדה. ראו לדוגמה בג'ץ 38379-12-24 **פלונת נ' בית הדין השרעי לערעורים ירושלים**, פס' 12 לפסק הדין (פורסם במאגר ממוחשב 23.2.25).

56 הטיה (או הפליה) אלגוריתמית היא מצב שבו אלגוריתם מזהה, במישור או בעקיפין, מידע ביחס למשתני רקע, כגון מגדר, גזע או נטייה מינית, ומפיק תוצאה מוטות בהינתן משתנים אלה. ראו רועי גולדשמידט **אפליה אלגוריתמית שמערכות המבוססות על בינה מלאכותית** (הכנסת מרכז המחקר ומידע לתשובה 2020). בעניינינו הטיה אלגוריתמית עשויה להתרחש בהינתן משתנים הקשורים במאפייניה של הרשות המקומית, לדוגמה מיקום גיאוגרפי, המצב הפיננסי שלה או המדד החברתי-כלכלי שאליו היא משתייכת.

57 כך לדוגמה, המודל עלול להצביע על מתאם שגוי בין דוחות ביקורת בנושאים שונים לחלוטין או לייחס ליקויים הנוגעים לרשות מקומית אחת לרשות מקומית אחרת. ראו גם Nelson F. Liu et al, Lost in the Middle: How Language Models Use Long Contexts, 12 Transactions of the Association for Computational Linguistics, Volume 157 (2024).

58 יחידות המידה של אורך הפלט והקלט הן לא מספר המילים אלא מספר הטוקנים (Token). טוקן הוא יחידת טקסט טכנית (שבר מילים, מילה שלמה או סימן פיסוק). קיים קשר ישר בין מספר הטוקנים ובין משך הזמן שנדרש מהמודל למתן התשובה. ככל שמספר הטוקנים גדול יותר, כך משך זמן ההמתנה לתשובה מתארך. ראו: Yuqing Yang, Lei Jiao & Yuedong Xu, A Queueing Theoretic Perspective on Low-Latency LLM Inference with Variable Token Length, Arxiv (December 2025) [ARXIV:2407.05347](https://arxiv.org/abs/2407.05347) [CS.NI].

59 כך לדוגמה, המודל עלול להצביע על מתאם שגוי בין מיקום גיאוגרפי של רשות מקומית מסוימת ובין דפוס ליקויים מסוים. ראו Marco J. Van Der Wal, The Marco-Effect: Functional Apophenia in Human-AI Interaction and the Asymmetry of Meaning-Bearing (2025) <https://philarchive.org/archive/VANFAI-3>.

60 לדוגמה, אם ישנה הטיה במאפייניה של הרשויות המקומיות הנכללות בביקורת, המודל עלול לזהות אותן באופן שגוי עם סיכון גבוה להימצאות ליקויים בהתנהלותן. הטיה אלגוריתמית עלולה להתרחש עקב הטיות במידע האימון או בשל בעיות בעיצוב המערכת.

61 ראו שור-עופרי "מעבר לשקיפות", לעיל ה"ש 52.

בדוחות הביקורת כרקע לביקורת או לממצאים שהועלו בה. אשר למגבלות המובנות במודל בהקשר של תהליך ניתוח המידע, בבדיקת ההיתכנות הראשונית לא נמצאו מקרים שבהם המודל זיהה דפוסים ומתאם (Correlation) שגויים. הסבר אפשרי לכך הוא שתהליך המטא-ביקורת, הכולל ניתוח רפלקטיבי של תהליך הביקורת עצמו, נועד מעצם טיבו לחשוף דפוסים והטיות בעבודת הביקורת. עם זאת, תופעות אלו מורכבות יותר לזיהוי לעומת השמטה של דוחות ביקורת, ומובן שנדרשות בדיקות נוספות כדי לאשר או לשלול את הימצאותן. אשר לתופעה של הטיות אלגוריתמיות, ניכר שתוצאות הניתוח הדהדו דפוסים קיימים בבחירת נושאי הביקורת. הם התמקדו בנושאי ביקורת שניתוח רפלקטיבי של תהליך הביקורת עצמו העלה שנבדקים בתדירות הגבוהה ביותר: תכנון ובנייה, רכש והתקשרויות, רווחה ושירותים חברתיים וכן מערכות מידע וסייבר.

סיכונים בשלב התיקוף של מהימנות התוצאות ושלמות

בינה מלאכותית, ובפרט מודלי שפה גדולים, נתפסת כ"קופסה שחורה" - מערכת המקבלת קלט (Input) ומפיקה פלט (Output), בעוד שתהליכי החשיבה שלה אינם מובנים במלואם ואי אפשר להתחקות אחריהם⁶². בהקשר של תהליך המטא-ביקורת המשמעות היא שגם אם הדוחות עצמם גלויים למשתמשים, תהליך הניתוח המתכלל, הסקת המסקנות והפקת תובנות-העל אינו שקוף להם, ויכולתם להבין תהליך זה במלואו ולהתחקות אחריו היא מוגבלת. היכולת להבין את תהליכי החשיבה של המודל ולהתחקות אחריהם קרויה הסבריות (Explainability). הסבריות ושקיפות בתהליכי קבלת החלטות הן שלב

ומתאפיינים בחוסר אחידות לשונית⁶² ומתודולוגית. דוחות הביקורת בשלטון המקומי, בפרט הדוחות המערכתיים מהשנים האחרונות, מתאפיינים באזכור של גופים מבוקרים רבים ברמת העמקה משתנה⁶³ (Data Noise). מאפיינים אלה מציבים אתגר רב למודלי שפה גדולים, בייחוד כאשר מתבצע חיפוש של התאמה מילולית מדויקת (String Match), אז הם עלולים להתקשות בזיהוי קשרים בין נושאים ובין מונחים ההכרחיים ליישום המטא-ביקורת. זוהי המחשה של העיקרון שלפיו קיים קשר ישר בין איכות הקלט לאיכות הפלט, וככל שאיכות הקלט נמוכה יותר, כך כוחות גם איכות הפלט (Garbage In, Garbage Out; GIGO).

בענייננו, בדיקת ההיתכנות הראשונית איששה את החשש מהשמטה בשוגג של דוחות ביקורת. שימוש בהנחות מדורגות ובשיטת ה-Slicing⁶⁴ הצביע על כך שההשמטה נגרמה בחלקה לפחות בשל המגבלות המובנות במודל. בדיקת ההיתכנות כללה שימוש בטכניקות נוספות לשיפור תהליך השליפה, כגון הוספת דרישה לשקיפות מלאה לגבי הדוחות שהושמטו ולגבי הדוחות שלא הושמטו בחלוקה נושאית ולפי שנת הפרסום, הכללת מנגנוני בקרה עצמית המחייבים את המודל לבדוק את תוצאותיו (Self-Consistency) והוספת דוגמאות מנחות לרמת הפירוט הנדרשת (Few-Shot)⁶⁵. אולם גם התוצאה המלאה ביותר שהתקבלה לא הייתה שלמה, וגם בה נמצאו השמטות.

בדיקת ההיתכנות הראשונית איששה גם את החשש משגיאות וטעויות הקשורות באיכות הקלט. בחלק מן התוצאות נמצא כי המודל זיהה ממצאים מתוך דוחות הביקורת כממצאיה של ביקורת המדינה, אף על פי שהממצאים הועלו בביקורות פנימיות של הגופים המבוקרים עצמם או בביקורות חיצוניות אחרות והובאו

62 לדוגמה, השימוש במונח "שקיפות" (לעומת "פרסום מידע"); שימוש במונחים של "ממשליות" (לעומת "סדירות", "ניהול סיכונים", "חוסן", "תכנון אסטרטגי") או "יעילות" (לעומת "מועילות וחסכון").

63 במקרים אלה שמם של הגופים המבוקרים אינו מופיע בהכרח בכותרת או בתת-פרק ייעודי, אלא נבלע בתוך קטעי טקסט עתירי נתונים או בקטע טקסט בודד במסמך של עשרות עמודים.

64 לדוגמה, לצורך זיהוי דפוס ליקויים חוזר ובעיות שורש ניתנה למודל הנחיה מפורשת להציג את תוצרי המיפוי הראשוני ולבקש את אישורו של המשתמש להתקדם לשלב הניתוח המתכלל. באופן זה נוצר דירוג בהנחיה (Chain of Thought), וללא אישור מפורש של המשתמש (In-Context Validation), אחרי שבוצעה בקרה אנושית על תוצרי המיפוי הראשוני, לא היה יכול המודל להתקדם לשלב הניתוח המתכלל.

65 לדוגמה, לצורך זיהוי דפוס ליקויים חוזר ובעיות שורש ניתנה למודל הנחיה מפורשת לבצע את המיפוי הראשוני של הדוחות ולהפיק ציר זמן (Timeline) בפורמט טבלה, כדלהלן: בעמודה הראשונה מספר סידורי; בעמודה השנייה שם הדוח; בעמודה השלישית סוג הביקורת (עומק/אחר); בעמודה הרביעית שיוך נושאי; בעמודה החמישית ממצא עיקרי. פורמט הטבלה והמידע שנכלל בה נועדו להקל את הבקרה האנושית על תוצרי המיפוי הראשוני, ובפרט כדי לזהות השמטה של מידע בשוגג. טכניקות נוספות שאפשר להסתייע בהן לשיפור תהליך השליפה הן טיוב הנתונים (Data Hygiene) ומיפוי מושגי (Ontology Mapping). טכניקות אלה מורכבות יותר לביצוע, ולא היה אפשר אותן בשלב בדיקת ההיתכנות הראשונית.

66 בהקשר זה נהוג גם לדבר על המונח Interpretability או Mechanistic Interpretability, כלומר היכולת להבין ולהסביר את תהליכי הפעולה הפנימיים של מודלי שפה גדולים. ראו (21.5.24) Anthropic, Mapping the Mind of a Large Language Model www.anthropic.com/news/mapping-mind-language-model

תהיה זו שמקבלת את ההחלטה בעצמה, אלא תשמש רק כלי תומך בתהליך קבלתה.

סיכום

במאמר זה הוצגה המטא-ביקורת כשלב נוסף ומתקדם בהתפתחות תחום ביקורת המדינה. שורשיה הרעיוניים טמונים בעולמות המטא-אנליזה והמטא-הערכה, המותאמים לעולם ביקורת המדינה. עוד יותר ממודלים קיימים בעולם, מודל המטא-ביקורת המוצע משלב בחינה מתכללת של ממצאי דוחות ביקורת קודמים עם ניתוח רפלקטיבי של תהליך הביקורת עצמו, וכן חילוץ תובנות-על (Meta-insights) שבאפשרותן להעצים את ערכה של הביקורת למוסד הביקורת עצמו, לקובעי המדיניות ולמקבלי ההחלטות בגופים העומדים לביקורת. תחום מתאים במיוחד ליישומה של הפרדיגמה המוצעת הוא ביקורת השלטון המקומי בישראל, לנוכח המבנה המבוזר והדו-רובדי המאפיין את השלטון המקומי ורעש הנתונים המוגבר המאפיין את הביקורת עליו.

בדיקת ההיתכנות הראשונית שבוצעה באמצעות NotebookLM על 260 דוחות ביקורת בשלטון המקומי (2010 - 2026) הדגימה את היכולת של כלי בינה מלאכותית זה לחלץ תובנות-על משלושה סוגים: (א) זיהוי דפוס ליקויים חוזר ובעיות שורש בהתנהלותן של הרשויות המקומיות; (ב) זיהוי כשלי אסדרה וכשלים מערכתיים ("ריק רגולטורי", "אשליית הקדמה הטכנולוגית", "איים של מידע"); (ג) ניתוח רפלקטיבי של דפוסי בחירת הנושאים והגופים המבוקרים. הבדיקה העלתה כי הנחיות מדורגות ובקרה משולבת - אנושית וממוחשבת - מפחיתות השמטות וטעויות, אך גם הדגישה מגבלות מובנות של מודלי שפה גדולים, ובהן תופעת "האובדן באמצע" והטיות אלגוריתמיות.

המאמר ממחיש כי השלם גדול מסך חלקיו - ערכו של גוף הידע שנצבר במשרד מבקר המדינה עולה על ערכם של דוחות הביקורת הבודדים. המטא-ביקורת, כתפיסת עבודה מובנית המשלבת כלי בינה מלאכותית לצד אחריות אנושית, מהווה שינוי פרדיגמטי המאפשר ביקורת על הביקורת והעצמת תרומתה של ביקורת המדינה לניהול הסיכונים והתכנון האסטרטגי של הגופים המבוקרים.

מקדים והכרחי ביצירת אחריותיות (Accountability) של מקבלי ההחלטות בכלל ושימוש בכלי בינה מלאכותית בפרט. בהיעדרן של הסבריות ושקיפות אין אפשרות לבצע בקרה אנושית מלאה, וללא בקרה אנושית אין אפשרות לאתר ולזהות השמטה של מידע, שגיאות, הזיות או הטיות בתוצאות הניתוח שהתקבלו⁶⁷. בלא הסבריות ושקיפות בתהליך, הבינה המלאכותית אינה כלי התומך בהחלטה שמתקבלת בהפעלת שיקול דעת אנושי, אלא הופכת למקבלת ההחלטה בעצמה⁶⁸.

יישום המטא-ביקורת בכלי בינה מלאכותית והטמעת תוצריה בתהליכי העבודה הפנימיים של משרד מבקר המדינה מחייבים, אם כן, קביעה של מסגרת עבודה הכוללת תהליך בקרה אנושי קפדני ומחמיר. על מסגרת העבודה לכלול בדיקה אנושית של התובנה המתקבלת ותיקוף של מהימנותה ושלמותה: בדיקה אם התובנה מעוגנת במקורות המידע שהועלו למודל (Grounding); בדיקה אם לא הושמטו מקורות מידע רלוונטיים; וכן תיעוד מפורט של אופן התיקוף של התובנה וההחלטה אם לקבלה או לדחותה. אפשר לשלב בתהליך התיקוף של התובנה שהתקבלה שימוש במודלי שפה גדולים אחרים, אך ככלי תיקוף משלים בלבד. טכניקות כמו ניסוח הנחיות מדורגות (Chain of Thought) ושימוש בשיטת ה-Slicing, המסייעות בצמצום השמטת מידע בשוגג, יכולות לסייע גם הן בהגברת ההסבריות של המודל ובהתמודדות עם בעיית הקופסה השחורה.

ריכוז הסיכונים והאתגרים העיקריים ביישום מטא-ביקורת בכלי בינה מלאכותית, בחלוקה לשלוש קטגוריות המשקפות את שלבי התהליך, ודרכי ההתמודדות עימם, מוצג ב שלהלן.

יכולותיהם הטכנולוגיות המתקדמות של מודלי שפה גדולים בעלי חלון הקשר רחב מאפשרות ניתוח מובנה של מאות מסמכים במהירות גבוהה, פעולה שאי אפשר ליישם באמצעים אנושיים בלבד. לצד היתרונות הברורים הגלומים בשימוש בכלי בינה מלאכותית ליישום המטא-ביקורת, הוא אינו חף מהקשיים המובנים בעבודה עם כלי הבינה המלאכותית הקיימים. אומנם קשיים אלה עשויים להצטמצם עם הזמן בעקבות השיפור ביכולותיהם הטכנולוגיות של המודלים מחד גיסא והבקרה האנושית עליהם מאידך גיסא, אך יש להניח כי הם לא ייעלמו לחלוטין. הם מחייבים אפוא שילוב של מסגרת עבודה זהירה ואחראית, מסגרת שתבטיח כי הבינה המלאכותית לא

67 גדי פרל ותהילה שוורץ אלטשולר "מודל ליצירת שקיפות אלגוריתמית" **הצעה לסדר** 21 47 (2022).

68 וכשם שנאמר לפני למעלה מחמישה עשורים על יכולותיו של המחשב: "המחשב פוטר [על-כן] את האדם מהחישוב ולא מהחשיבה". ראו עמיקם רייכמן "עיבוד נתונים אוטומטי" **מערכות** 34 231 (1973).

תרשים ה: הסיכונים והאתגרים העיקריים ביישום מטא-ביקורת
בבלי בינה מלאכותית ומסגרת העבודה המוצעת להתמודדות עימם



הבינה המלאכותית בשירות הביקורת הכלכלית: ניתוח אוטומטי של נתונים כלכליים

האלה אבו-חלה ואופיר מזרחי

ד"ר האלה אבו-חלה היא מנהלת ביקורת באגף לביקורת תחום האנרגיה בחטיבה לביקורת תחומי הכלכלה והתשתיות הלאומיות שבמשרד מבקר המדינה ונציב תלונות הציבור.

מר אופיר מזרחי הוא אחראי לפיתוח למידה באגף הלמידה והפיתוח הארגוני שבמשרד מבקר המדינה ונציב תלונות הציבור.

המאמר בקצרה

- שילוב כלי בינה מלאכותית בביקורת הפיננסית עשוי לחולל מהפכה: הוא מאפשר עיבוד יעיל של נתוני עתק ואיתור מהיר של חריגות וסיכונים מתוך דוחות כספיים, ובכך לחסוך זמן ולייעל את תהליך הביקורת.
- מקרה בוחן בעניין ביקורת פיננסית על חברה ממשלתית, הראה כי כלי בינה מלאכותית הצליחו לזהות ליקויים ואף להמליץ לבצע ביקורת מעקב בהתבסס על השוואה של נתוני הדוחות הכספיים המעודכנים אל מול ממצאי הביקורת המקורית.
- אולם תוצאות הניתוח שנערך בכלי בינה מלאכותית נשען ברובו על נתונים היסטוריים בלבד, ואילו בדוח הביקורת גם מחברים את הנתונים ה"יבשים" להקשרים עסקיים, תכנוניים ואסטרטגיים, דבר המאפשר לעורכי הדוח להצביע על סיכונים אסטרטגיים שלא קיבלו ביטוי בדוחות הכספיים ושלא זוהו על ידי כלי ה-AI.
- המסקנה היא שהבינה המלאכותית משמשת כמכפיל כוח למבקרים, אך אינה יכולה להחליף את שיקול הדעת האנושי הנדרש להבנת התמונה הכלכלית המלאה בביקורת.

מבוא | ביקורת פיננסית | בינה מלאכותית וביקורת | שימוש בכלי בינה מלאכותית ליעול ביקורת כלכלית | סיכונים בשימוש בכלי בינה מלאכותית | מתודולוגיה לביצוע ביקורת פיננסית באמצעות כלי בינה מלאכותית של חברת גוגל | השימוש בכלי בינה מלאכותית בביקורת פיננסית על חברה ממשלתית ציבורית | המלצות הבינה המלאכותית לביצוע ביקורת מעקב | סיכום ומסקנות

מטרתו של מאמר זה היא לבחון את הפוטנציאל הגלום בשימוש בכלי AI לצורך ייעול תהליכי הביקורת הפיננסית. במאמר מוצגת מתודולוגיה לשימוש בכלים אלו בניתוח דוחות כספיים ובהשוואה לממצאי ביקורת מסורתיים, במטרה לבחון את תרומתם לשיפור יכולת הזיהוי של כיווני ביקורת, ליקויים וסיכונים פיננסיים. לצורך המחשה אמפירית המתודולוגיה מודגמת באמצעות ניתוח דוחות כספיים של חברה ממשלתית ציבורית. זאת ועוד, באמצעות ניתוח זה יודגם כיצד שילוב של מומחיות מקצועית של צוותי ביקורת ויכולות עיבוד נתונים מתקדמות של בינה מלאכותית עשוי לשפר את איכות הביקורת מזה ולהרחיב את היקף הבדיקה ועומקה מזה.

ביקורת פיננסית

על פי חוק יסוד: מבקר המדינה, מתפקידו של מבקר המדינה לבחון את חוקיות הפעולות, טוהר המידות, הניהול התקין, היעילות והחיסכון של הגופים המבוקרים, וכל עניין אחר שהוא רואה בו צורך. חוק מבקר המדינה, התשי"ח-1958 [נוסח משולב], קובע כי "במסגרת תפקידו יבדוק המבקר במידת הצורך – אם כל הוצאה הוצאה בתחום ההקצבה החוקית ולמטרה שנועדה לה; אם ההכנסות נתקבלו לפי החוק ומותרות לפי החוק; אם יש לכל הוצאה או הכנסה מסמך מספיק; אם כל פעולה שבתחום בקרתו נעשתה בהתאם לחוק ובידי המוסמך לעשותה; אם הנהלת החשבונות, עריכת המאזנים, בקורת הקופה והמלאי ושיטת ניהול המסמכים הן יעילות; אם אופן החזקת הכספים ושמירת הרכוש מניח את הדעת; אם מצב הקופות ומלאי הרכוש מתאים לחשבונות; ואם הגופים המבוקרים ... נהגו בחסכון וביעילות ובטוהר המידות". אם כן, ביקורת פיננסית היא כלי מרכזי לבדיקת היעילות והחיסכון בגופים המבוקרים.

ביקורת פיננסית מוגדרת כתהליך שיטתי, בלתי תלוי ומבוסס ראיות, שמטרתו לחוות דעה אם הדוחות הכספיים משקפים באופן נאות, מכל הבחינות המהותיות, את מצבה הכספי של הישות בהתאם למסגרת הדיווח הרלוונטית¹.

ארגון הביקורת הבין-לאומי (International Organization of Supreme Audit Institutions; INTOSAI) קבע עקרונות וסטנדרטים בין-לאומיים לביקורת פיננסית. לפי הארגון, הביקורת הפיננסית היא מרכיב מרכזי בעבודתם של

ביקורת פיננסית היא בדיקה בלתי תלויה שמטרתה לבחון אם המידע הכספי של הגוף המבוקר מוצג בהתאם למסגרת המאסדרת (הרגולטורית) שחלה עליו (ISSAI 100), ואם המידע הזה משקף באופן נאות את מצבו הכלכלי של הגוף המבוקר.

הביקורת הפיננסית היא נדבך מרכזי במנגנוני הפיקוח והבקרה על פעילותם של גופים ציבוריים. באמצעות בחינת הדוחות הכספיים, מערכות הבקרה הפנימית ואופן ניהול המשאבים, מבקשת הביקורת להבטיח שימוש תקין, יעיל ושקוף בכספי הציבור, לחזק את הנשיאה באחריות (Accountability) של הגופים המבוקרים ולהגביר את אמון הציבור במוסדות השלטון. בהתאם לכך, מוסדות ביקורת המדינה ברחבי העולם רואים בביקורת הפיננסית כלי מרכזי לאיתור סיכונים פיננסיים, לזיהוי ליקויים בממשל התאגידי ולהבטחת ניהול תקין של משאבים ציבוריים.

בעשורים האחרונים חלה עלייה חדה בהיקף הנתונים הפיננסיים ובמורכבות הסביבה הכלכלית והמאסדרת שבה פועלים ארגונים. התרחבות מאגרי המידע והמעבר לעולמות של נתוני עֵתֶק (Big Data) יוצרים קושי בניתוח פיננסי מעמיק, זאת לנוכח הפער בין יכולות הביקורת המסורתיות, המבוססות על איסוף ועיבוד נתונים בדרכים מסורתיות², להיקף ולרמת המורכבות של המידע הקיים בידי הארגונים. לעיתים מגבלות של זמן, של משאבים ושל יכולת אנושית לעבד כמויות מידע עצומות מצמצמות את היקף הבדיקה ומגבילות את יכולת המבקרים לזהות דפוסים חריגים, אנומליות פיננסיות וסיכונים מערכתיים.

התפתחויות טכנולוגיות בתחום הבינה המלאכותית (Artificial Intelligence; AI), ובעיקר למידת מכונה, מודלים של ניבוי ובינה מלאכותית יוצרת (Generative AI) פותחות אפשרויות חדשות לשיפור תהליכי הביקורת הפיננסית. כלים אלה מסוגלים לעבד כמויות עצומות של נתונים בזמן קצר, לזהות דפוסים מורכבים, לבצע ניתוחים השוואתיים מתקדמים ולהפיק תובנות המסייעות באיתור מוקדם של סיכונים פיננסיים. לפיכך, שילוב מערכות בינה מלאכותית בעבודת הביקורת בכלל וביקורת הפיננסית בפרט עשוי לשנות במידה רבה את האופן שבו מתבצע הניתוח הכלכלי, ממלאכה ידנית וממושכת לניתוח אנליטי מתקדם הנתמך בטכנולוגיה.

1 International Organization of Supreme Audit Institution (INTOSAI), ISSAI 100 - Fundamental Principles of Public-sector Auditing (2019), https://www.intosai.org/fileadmin/downloads/documents/open_access/ISSAI_100_to_400/issai_100/ISSAI_100_EN.pdf.

2 בעשורים האחרונים נעשה שימוש בכלים ממוחשבים שהולכים ומשתכללים. עם זאת, לצורך שימוש בכלים אלה לרוב נדרש לאסוף את הנתונים, בייחוד כשמדובר בנתונים ממקורות שונים, ולהעלותם על אותו גיליון כדי שיהיה ניתן לעבדם.

3 INTOSAI, ISSAI 100, לעיל ה"ש 1.

נוכח חשיבותה, משרד מבקר המדינה מבצע מדי פעם בפעם ביקורת פיננסית המתמקדת בבחינת הסיכונים, המגמות והאתגרים העולים מניתוח הדוחות הכספיים של הגופים המבוקרים, בחינה של מנגנוני הבקרה הפיננסיים ובחינה כלכלית של סוגיות שונות העולות מהם.

היועץ הכלכלי במשרד מבקר המדינה פיתח בשנים האחרונות מתודולוגיה מומלצת לביצוע ביקורת פיננסית. על פי מתודולוגיה זו, ביקורת פיננסית נשענת על שלוש אבני יסוד מרכזיות: ניתוח של היחסים הפיננסיים והכלכליים; ניתוח הסיבות לשינויים ביחסים ובמשתנים הפיננסיים והכלכליים; אמירת ביקורת⁸. כדי לקבל תמונה פיננסית רחבה על פעילות הגוף המבוקר, מומלץ כי הביקורת תכלול ארבעה היבטים: (א) ניתוח של הדוחות הכספיים המבוקרים; (ב) ניתוח של היחסים הפיננסיים והכלכליים⁹; (ג) בחינת העבודה של רואה החשבון המבקר; (ד) סקירה של סוגיות בממשל התאגידי בגוף המבוקר¹⁰.

בחינת היחסים הפיננסיים העיקריים של הארגון (הגוף המבוקר) היא שלב חשוב בניתוח הדוחות הכספיים שלו, ויש להביא בחשבון שקיים מספר רב של יחסים פיננסיים נפוצים. הנתונים המוצגים ביחסים הפיננסיים נאספים בשיטות ובאופן טכני¹¹ מתוך הדוחות הכספיים ולעיתים גם ממידע נוסף הנכלל בפרק התיאור העסקי המצורף לדוחות הפיננסיים. כל המידע והנתונים מאפשרים תיאור וניתוח של היבטים שונים בפעילות הארגון. ניתוח היחסים הפיננסיים של הארגון כולל בחינה בשתי רמות:

מוסדות ביקורת המדינה לצורך פיקוח על כספי הציבור, ומטרתה העיקרית היא לעמוד על הסיכונים הפיננסיים או על האינדיקציות לקיומם בגוף המבוקר, לרבות בדיקת אמינות הדוחות הכספיים של הגופים הציבוריים ושמירה על שימוש תקין בכספי ציבור, לאתר נקודות תרופה בפעילותו של הגוף המבוקר ולבחון את מנגנוני הבקרה הפיננסיים בו⁴. בהקשר הציבורי, לביקורת הפיננסית חשיבות רבה בחיזוק האחריותיות השלטונית ובהבטחת השימוש התקין בכספי הציבור⁵, בקידום ממשל תקין ושקיפות ציבורית ובמניעת שחיתות והונאה. משרד מבקר המדינה האמריקאי⁶ ציין אף הוא כי ביקורת פיננסית במגזר הציבורי עשויה להגביר את אמון הציבור בשל היותה ביקורת בלתי תלויה הבוחנת את המידע הפיננסי המדווח, וכי היא תורמת לצמצום הסיכונים בשימוש בכספי ציבור.

לביקורת הפיננסית יתרונות רבים: היא תורמת לחיזוק השקיפות, לאימות נכונות הדיווחים הכספיים ולמניעת הצגת מידע מוטעה; מעצימה את הנשיאה באחריות של הגופים הציבוריים, שכן גופים אלה נוטים לפעול בזירות רבה יותר כשהם יודעים שהם נתונים לביקורת פיננסית; מסייעת למניעת הונאות באמצעות זיהוי אנומליות או חריגות בנתונים ולזיהוי חולשות בבקרה הפנימית ונותנת המלצות לתיקון; מגבירה את אמון הציבור, שכן הביקורת נעשית על ידי גוף חיצוני בלתי תלוי; משפרת את תהליכי קבלת ההחלטות ואת יעילות הפעולות בארגון⁷.

4 ירון פישמן **ביקורת ביצועים בעלת היבטים פיננסיים וכלכליים במיקוד על חברות ממשלתיות** (לשכת מבקר במדינה ונציב תלונות הציבור 2024), <https://media.mevaker.gov.il/mevaker/media/oyrh3jef/financial-audit-1.pdf>.

5 Miranda Misini & Krisela Ngjela, Strengthening Public Fund Governance through Financial Audits: Insights from the Albania Supreme Audit Institution's Practice, 52(3) Int'l J. of Government Auditing 20 (2025).

6 U.S. Government Accountability Office (GAO), Financial Audit: FY 2024 and FY 2023 Consolidate Financial Statement of the U.S. Government (GAO-25-107421) (2025), <https://www.gao.gov/products/gao-25-107421>.

7 Misini & Ngjela, Strengthening Public, לעיל ה"ש 5.

8 פישמן, **ביקורת ביצועים**, לעיל ה"ש 4.

9 לרבות יחסי נזילות המשקפים את יכולת הגוף המבוקר לעמוד בהתחייבויותיו בזמן קצר, כגון היחס השוטף, היחס המהיר, יחס תזרים מפעילות שוטפת להתחייבויות שוטפות; המינוף הפיננסי שמשקף את מבנה ההון של התאגיד; יחסי כיסוי חוב העוסקים ביכולת התאגיד לשרת את החוב הפיננסי שלו, כגון יחס החוב הפיננסי נטו לרווח לפני ריבית, מיסים, פחת והפחתות (EBITDA), יחס החוב הפיננסי נטו לתזרים מפעילות שוטפת (FFO); יחס כיסוי הריבית שבוחר את יכולת התאגיד לעמוד בנטל תשלומי הריבית השוטפת שלו; רווחיות, כגון שיעור הרווח הגולמי מהמחזור, יחס EBITDA להכנסות, התשואה על הנכסים ועוד.

10 הממשל התאגידי הוא אוסף מנגנוני פיקוח ובקרה שמטרתם לתרום לניהול התקין של התאגיד, לשמירה על הוראות החוק ולשמירה על האתיקה של הארגון העסקי ועל נאותות הדיווח הכספי, כל זאת לצורך הגנה על בעלי העניין בתאגיד הציבורי. דוגמאות לסוגיות בממשל תאגידי שניתן לבדוק: אישור הדירקטוריון של הגוף המבוקר, אישור דוחות כספיים, דיווח על אפקטיביות הבקרה הפנימית, ניהול סיכונים פיננסיים בתאגיד, פרסום דוחות לציבור ועוד.

11 פעולה טכנית הכרוכה בחילוץ הנתונים מתוך הדוחות והצבתם בנוסחה.

בינה מלאכותית וביקורת

מושגי יסוד בבינה מלאכותית

טכנולוגיית הבינה המלאכותית מאפשרת למערכות ממוחשבות לבצע או לסייע בביצוע מגוון רחב של משימות שבאופן מסורתי דרשו עבודה אינטלקטואלית אנושית. משימות אלה כוללות לדוגמה ניתוח נתונים מורכבים, בניית מודלים וניתוח תהליכים של קבלת החלטות¹⁴. בעבר היו מערכות בינה מלאכותית מערכות סטטיות שפעלו על פי כללים מוגדרים מראש במטרה לסווג מידע ולמצוא סדר בנתונים קיימים. אולם בשנים האחרונות חלו התפתחויות מרחיקות לכת ביכולות הביצוע של הבינה המלאכותית, כפי שמתואר להלן בתרשים 1.

בשלב הראשון עברה הטכנולוגיה לשימוש באלגוריתמים של למידת מכונה (Machine Learning), דבר שאיפשר למערכות לנתח כמויות עצומות של נתונים, לזהות דפוסים מורכבים ולבצע התאמות לתנאים משתנים בדיוק וביעילות¹⁵.

התפתחות בולטת נוספת של טכנולוגיה זו היא הבינה המלאכותית היוצרת (Generative AI), המסוגלת ליצור תכנים חדשים כמו טקסט, תמונות ואודיו באמצעות למידה ממאגרי נתונים קיימים. שלא כמודלים מסורתיים, המתמקדים בסיווג מידע, בינה מלאכותית יוצרת מפיקה תכנים חדשים לחלוטין בתבניות חדשות ומגוונות: טקסטים, תמונות, שמע, וידאו וקוד מחשב¹⁶. חלק מרכזי בטכנולוגיה זו נשען על מודלי שפה גדולים (Large Language Models; LLMs), המבוססים על ארכיטקטורה ייחודית המאפשרת להם להבין מחרוזות של טקסט. מודלים אלו לומדים כמויות טקסט אדירות כדי להעריך, לזהות ולחזות תבניות והקשרים לשוניים, ודבר זה מאפשר להם להבין את ההקשר של השפה הטבעית ולייצר טקסטים ברמה לוגית וקוהרנטית גבוהה¹⁷ הנראים כאילו נוצרו בידי בני אדם.

1. **בחינה לאורך זמן:** התפתחות היחסים הפיננסיים של הארגון לאורך זמן (לדוגמה, כמה שנים או כמה רבעונים), כדי ללמוד על מגמות ועל התפתחויות שחלו במצבו הכספי ובתוצאות פעילותו.

2. **בחינה השוואתית בנקודת זמן הכוללת שני סוגי בחינה עיקריים:**

א. השוואת היחסים הפיננסיים של הארגון ליחסים הפיננסיים בתאגידי השוואה הפועלים באותו ענף (Benchmark);

ב. השוואת היחסים הפיננסיים של ארגון שהוא מונופול מפוקח ליחסים מקובלים שקבע המאסדר¹².

להשוואות אלה ערך מוסף רב. הן יכולות להצביע על מגמות ארוכות טווח ועל שינויים שחלו בתמונת המצב העסקית והפיננסית של הגוף המבוקר, להעריך את מצבו בהשוואה לגופים דומים ולאפשר מדידה של תוצאותיו העסקיות של הגוף המבוקר. אולם כדי שהניתוח המשווה יהיה בעל תוקף ובעל מהימנות גבוהה, נדרש לבצע השוואות בטווח זמן ארוך מספיק ולהשוות לכמה גופים בעלי מאפייני פעילות דומים. לביקורת המסורתית, הנשענת על איסוף ועיבוד ידיניים של נתונים פיננסיים, יש מגבלות הנובעות, בין היתר, ממגבלות זמן ומיכולת מוגבלת של צוות הביקורת לאסוף כמויות מידע עצומות (נתוני עתק) בפרקי זמן סבירים וברמת דיוק גבוהה.

הפער בין יכולות הביקורת המסורתיות ובין מורכבות המידע בסביבה המבוקרת בכלל ומאגרי הנתונים הפיננסיים בפרט, הביא כאמור לפיתוח טכנולוגיות וכלים אנליטיים מתקדמים לניתוח המידע ועיבודו. לאור ההתפתחות הטכנולוגית המהירה, הכרחי שבעבודת הביקורת יאומצו טכנולוגיות חדשניות אלה לניתוח נתונים ועיבודם, כגון כלי AI ומודלים של למידת מכונה¹³.

12 פישמן, **ביקורת ביצועים**, לעיל ה"ש 4.

13. Tiare Rivera, Machine Learning Application for SAIs, 50(2) Int'l J. of Government Auditing 13 (2023). את היכולת לנתח כמויות עצומות של נתונים, לזהות דפוסים מורכבים ולבצע התאמות באופן דינמי לתנאים משתנים בדיוק וביעילות. עוד על כך ראו להלן ב"מושגי יסוד בבינה מלאכותית".

14. Mohammad Hosseini, Maya Murad & David B. Resnik, Benefits and Risks of Using AI Agents in Research, 56(1) The Hastings Center Report 13 (2026), <https://doi.org/10.1002/hast.70025>.

15. Wilhelmina Afua Addy et al., AI in Credit Scoring: A Comprehensive Review of Models and Predictive Analytics, 18(2) Global J. of Engineering and Tech. Advances 118 (2024), <https://doi.org/10.5281/ZENODO.10947394>.

16. Antoine Bordas et al., What is Generative in Generative Artificial Intelligence? A Design-Based Perspective, 35(4) Research in Engineering Design 427 (2024), <https://doi.org/10.1007/s00163-024-00441-x>.



התרשים עוצב באמצעות כלי הבינה המלאכותית Gemini.

שפה גדולים (LLM) מיועדים בעיקר למילוי משימה בודדת, ואילו סוכני בינה מלאכותית משלבים יכולות של תכנון, קבלת החלטות וביצוע פעולות בסביבה משתנה. חיבור זה מאפשר לסוכן לקבל משימות מורכבות, לפרק אותן לתת-שלבים ולבצע אותן באופן עצמאי. סוכנים אלו מסוגלים לתכנן ולהוציא לפועל מהלכים באופן הגיוני, לתאם בין משימות ולהקצות משאבים כדי להשלים באופן עצמאי שרשרת של משימות הקשורות זו בזו מראשיתן ועד סופן.¹⁹ ככל שכלי ה-AI הולכים ומתפתחים, כך יכולות איסוף הנתונים, עיבודם והסקת המסקנות מהם הולכות ומשתכללות. היום כלים אלה מסוגלים לעבד נתוני עתק (Big Data) בזמן אמת, לזהות מגמות ולהפוך תהליכים ידניים מְכֻרְכִים לפשוטים באמצעות אוטומציה של משימות ופעולות שגרתיות.²⁰ טכנולוגיית הבינה המלאכותית מציעה

סוג נוסף של מערכות מבוססות נתונים הוא מודלים לחיזוי (Predictive Models), המתמקדים בניתוח נתונים היסטוריים כדי לזהות מגמות ולנבא התנהגויות ותרשישים עתידיים. באמצעות שילוב אלגוריתמים מתקדמים, כגון עצי החלטה, רגרסיה ליניארית ורגרסיה לוגיסטית, מודלים אלו פועלים באופן דינמי ומתאימים את עצמם באופן רציף לדפוסים משתנים. כלים אלה משמשים לניתוח נתונים מורכבים, והם מסוגלים גם לשלב בזמן אמת מידע מכמה מקורות נתונים, ובכך לאפשר תהליכי קבלת החלטות מדויקים ומושכלים יותר, תוך ניהול סיכונים.¹⁸

השלב המתקדם ביותר בהתפתחות כלי ה-AI בא לידי ביטוי בסוכני בינה מלאכותית (AI Agents). אלה מערכות או תוכנות המסוגלות לבצע משימות באופן אוטונומי ללא צורך בפיקוח או בשליטה חיצונית של בני אדם. מודלי

18 Addy et al., AI in Credit Scoring, לעיל ה"ש 15.

19 Hosseini, Murad & Resnik, Benefits and Risks, לעיל ה"ש 14.

20 Addy et al., AI in credit scoring, לעיל ה"ש 15.

הפקת תובנות תומכות החלטה. לשימוש בכלים אלה בביקורת פיננסית יתרונות בולטים:

- **ניתוח כמויות מידע גדולות:** כלי נתוני עֶתֶק ולמידת מכונה מאפשרים לעבד מיליוני רשומות חשבונאיות ולאתר חריגות²⁴;

- **אוטומציה של פעולות שגרתיות ושיפור יעילות תפעולית:** אוטומציה של תהליכים חוזרים (Automation; RPA) מאפשרת להפחית את כוח האדם המיועד למשימות אלה, לקצר את משך מחזור הביקורת ולהקצות למבקרים יותר זמן לפעולות מורכבות הדורשות העמקה, חשיבה, שיקול דעת מקצועי והערכת סיכונים אסטרטגיים²⁵.

- **הרחבת היקף הבדיקה ושיפור איכות הראיות:** אלגוריתמים מתקדמים מאפשרים לנתח אוכלוסיות מלאות של נתונים במקום דגימות חלקיות²⁶. Vasarhelyi, Kogan & Tuttle²⁷ טענו שניטור רציף (Continuous Auditing) משפר את יכולת הזיהוי המוקדם של כשלים פיננסיים.

- **שיפור היכולת לגלות הונאות ולחשוף סיכונים מערכתיים:** כלי AI, ובראשם למידת מכונה, מגבירים את היכולת לגלות הונאות ומידע שגוי. Kokina & Davenport²⁸ הצביעו על כך שבכל הקשור לזיהוי הונאות, מודלים של למידת מכונה מדויקים יותר ממודלים סטטיסטיים מסורתיים, וכי יכולת זו בולטת במיוחד בסביבות עתירות עסקאות ונתונים.

- **ניתוח מסמכים טקסטואליים והסקת מסקנות מהם:** מערכות לעיבוד שפה טבעית (Natural Language

אפוא הזדמנויות רבות ובעלות ערך של ממש בתחומי הביקורת, ויש בה פוטנציאל לשינוי המקצוע מן היסוד בזכות היכולת לעבד ולנתח כמויות נתונים עצומות ולהפיק תובנות חדשות.

שימוש בכלי בינה מלאכותית ליעול ביקורת כלכלית

שימוש בכלי AI ובטכנולוגיות מידע מתקדמות בביקורת כלכלית מאפשר עיבוד נתוני עֶתֶק וניתוחם במהירות ובדיוק, בתהליך שיטתי ורציף, ודבר זה מסייע לביצוע עבודת ביקורת יעילה ואפקטיבית²¹. בכוחה של טכנולוגיית הבינה המלאכותית לשפר את יעילותה של הביקורת הכלכלית, עקב המעבר מביצוע משימות ידניות ושגרתיות, כגון סקירה, קריאה וניתוח של מספר רב של אסמכתאות, לניתוח נתונים מתקדם ולהפקת תובנות חדשות ומעמיקות.

המעבר לניתוח נתונים מתקדם מתבטא גם ביכולות של זיהוי דפוסים, סיווג תמונות וניתוח טקסט, המחליפות את בדיקת המסמכים המסורתית. מערכות לזיהוי תמונה מסוגלות לחלץ מידע מובנה ממסמכים סרוקים (לדוגמה, מסמכים בתבנית PDF), ובכך להרחיב את היקף הבדיקות ולאתר מגמות או חריגות המחייבות בדיקה מעמיקה. במישור הטקסטואלי אפשר, למשל, לנתח אסמכתאות רבות כדי לאתר ליקויים²².

D. Leocádio, L. Malheiro & J. Reis²³ זיהו ארבעה תחומים ליישום של כלי AI בביקורת פיננסית: זיהוי אנומליות, הערכת סיכונים הונאה, ניתוח טקסטואלי של דוחות כספיים

1. INTOSAI, ISSAI 100 ה"ש 1.

22 Nir Eisikovits, William C. Johnson & Ariel Markelevich, Should Accountants Be Afraid of AI? Risks and Opportunities of Incorporating Artificial Intelligence into Accounting and Auditing, 39(2) Accounting Horizons 117 (2025),

<https://doi.org/10.2308/HORIZONS-2023-042>.

23 Diogo Leocádio, Luis Malheiro & Joao Reis, Artificial Intelligence in Auditing: A Conceptual Framework for Auditing Practices, 14(10) Administrative Science 238 (2024), <https://www.mdpi.com/2076-3387/14/10/238>

24 Rivera, Machine Learning 13, לעיל ה"ש 13.

25 Julia Kokina & Thomas H. Davenport, The Emergence of Artificial Intelligence: How Automation is Changing Auditing, 14(1) J. of Emerging Tech. in Accounting 115 (2017),

https://www.researchgate.net/publication/315955305_The_Emergence_of_Artificial_Intelligence_How_Automation_is_Changing_Auditing.

26 Leocádio, Malheiro & Reis, Artificial Intelligence in Auditing 26, לעיל ה"ש 23.

27 Miklos A. Vasarhelyi, Alexander Kogan & Brad M. Tuttle, Big Data in Accounting: An Overview, 29(2) accounting horizons 381 (2015).

28 Kokina & Davenport, The Emergence of Artificial Intelligence 28, לעיל ה"ש 25.

שילוב של מערכות בינה מלאכותית בביקורת פיננסית עשוי להביא לשינוי חיובי בתפקידים המקצועי של המבקרים, רואי החשבון ואנשי הפיננסים. כשהטכנולוגיה מבצעת אוטומטית משימות שגרתיות, כגון בדיקת נתונים טכנית וחישוב יחסים פיננסיים, אנשי המקצוע מתפנים לעסוק בפעילויות חשיבה ברובד גבוה יותר, כגון שיפוט מקצועי מתקדם וייעוץ אסטרטגי, המספקות ערך מוסף לארגון. במילים אחרות, השימוש בכלי בינה מלאכותית והשינוי המתחולל בתחום הביקורת עשויים לשפר את איכות הביקורת ולאפשר התמקדות בהיבטים מורכבים יותר תוך הסתמכות על ניתוחים מתקדמים של כלי הבינה המלאכותית³¹.

סיכונים בשימוש בכלי בינה מלאכותית

לצד יתרונותיה הרבים של הבינה המלאכותית קיימים בה גם סיכונים, מקצתם כלליים ומקצתם ייחודיים לעבודת הביקורת. להלן לוח 1 הכולל סקירה של הסיכונים העיקריים בעבודת ביקורת המבוססת על כלי בינה מלאכותית³². מוסדות ביקורת אינם יכולים להתעלם מההתקדמות המהירה של הבינה המלאכותית ועליהם להכירה, הן לצורך הבנת עבודת הגופים המבוקרים המשתמשים בכלי AI

(Processing; NLP) ואלגוריתמים חכמים מאפשרים לנתח אוטומטית מסמכים כגון חוזים וכרטיסי הזמנות²⁹.

• **חיסכון בעלויות:** מאחר שהשימוש בכלי AI מייעל את עבודת הביקורת ומקצר את משך הביצוע שלה הוא מצמצם את עלויותיה³⁰.

יתרונות אלה מאפשרים למבקרים לבצע ניתוחים מורכבים מאוד, שבעבר לא היה ניתן לבצעם, ולהתמודד עם כמויות מידע גדולות ונתוני עֶתָק בפרקי זמן קצרים.

טכנולוגיה בולטת נוספת העשויה לשפר את עבודת הביקורת היא הבינה המלאכותית היוצרת, המבוססת על מודלים של עיבוד שפה טבעית, כגון Gemini-i ChatGPT, המסוגלים להבין, לנתח ולהפיק תכנים טקסטואליים ברמה גבוהה. יישומים אלו יכולים לשמש בזיהוי תובנות ובהעמקת תהליכי הביקורת המבוססים על הנתונים החשבונאיים הגולמיים של החברה. תחום מרתק נוסף המתפתח במהירות הוא מודלים לחיזוי המשמשים לזיהוי מגמות ולתמיכה בקבלת החלטות באמצעות חישובים אנליטיים אוטומטיים המבוססים על נתונים בתחום הביקורת. שימוש בכלים אלה מאפשר למבקרים להעריך את מצבה הפיננסי של החברה, לזהות נושאים החורגים מהציפיות ולהצביע על סיכונים.

לוח 1: סיכונים עיקריים בשימוש בכלי בינה מלאכותית בעבודת הביקורת

הנושא	תיאור הסיכונים
הגנת הפרטיות	מערכות בינה מלאכותית עשויות לפגוע בפרטיות עקב שימוש במידע רגיש לצורך אימון המודלים. שימוש בנתונים רגישים של ארגונים, כגון פרטים של לקוחות או של עובדים, עלול לדלוף ולפגוע בפרטיות.
דליפת מידע סודי, מסווג וקנייני	יש חשש שמערכות בינה מלאכותית, ובעיקר מודלי שפה, ישתמשו במידע רגיש שהוזן אליהן. מודלים אלה עלולים לשלב נתונים רגישים בתוצרים ולחשוף מידע עסקי או אישי. זאת ועוד, המידע המוזן בכלים של חברת OpenAI למשל הופך פעמים רבות לקניינה של מפתחת המודל, דבר שאינו מתקבל על הדעת בתעשיות מפוקחות המחייבות לסודיות ולפרטיות.

²⁹ Eisikovits, Johnson & Markelevich, Should Accountants Be Afraid of AI 29, לעיל ה"ש 22.

³⁰ Dolores Genaro-Moya, Antonio Manuel Lopez-Hernandez & Marila Godz, Artificial Intelligence and Public Sector Auditing: Challenges (and Opportunities for Supreme Audit Institutions 6(2) World 1 (2025).

³¹ Eisikovits, Johnson & Markelevich, Should Accountants Be Afraid of AI 31, לעיל ה"ש 22.

טיב הנתונים	איכות התוצרים של מערכות בינה מלאכותית תלויה באיכות הנתונים שהוזנו אליה. נתונים חסרים, שגויים או מוטים עלולים להוביל להסקת מסקנות שגויות.
הטיות אלגוריתמיות	אלגוריתמים מבוססים על נתוני עבר, ולכן הם עלולים לשקף הטיות הקיימות בנתונים. כשהנתונים מוטים או אינם מייצגים באופן מאוזן, התוצאות עשויות להיות מוטות ולהוביל למסקנות שגויות. במצב זה ייתכן שהמערכת לא תזהה ליקויי ביקורת חדשים שאינם תואמים לתבניות העבר המוכרות לה.
המצאת עובדות שקריות (הזיות)	במערכות בינה מלאכותית קיימת תופעה של יצירת מידע שאינו מבוסס על נתונים אמיתיים (הזיות). סיכון זה מדגיש שמערכות אלו אינן תחליף למומחיות אנושית, ותוצרי המערכת צריכים להיבחן בקפדנות.
קופסה שחורה (Black Box)	לעיתים קרובות אי אפשר לשחזר, להסביר או להבין כיצד אלגוריתם של בינה מלאכותית הגיע להחלטותיו או להמלצותיו. היעדר שקיפות מסוג זה מקשה על בחינת אופן הפעולה ועל הערכת מהימנות התוצרים. זאת ועוד, בעקבות ההסתמכות הגוברת על אלגוריתמים שאי אפשר להסביר את תוצאות פעולתם עשוי להישחק האמון המקצועי בקרב אנשי המקצוע.
הסתמכות יתר על המערכת	שימוש נרחב בכלי בינה מלאכותית עלול להפחית את פעולות הביקורת האנושיות המסורתיות ולהישען בעיקר על תוצרי המערכות האוטומטיות. מצב זה עלול לפגוע הן באיכות הביקורת הן בשיקול הדעת האנושי.
אסדרה וחסימים מוסדיים	תחום הבינה המלאכותית הולך ומתפתח, והמסגרת המאסדרת, בין היתר גם בתחום הפיננסי, עדיין מתגבשת. יישום כלים אלה במוסדות ביקורת מחייב פיתוח כללים מנחים ונהלים כדי למנוע שימוש בלתי מבוקר בכלים אלו ולהבטיח איכות מקצועית.

על גופים מבוקרים באמצעות אוטומציה של ניתוחים השוואתיים, חישוב יחסים פיננסיים ומודלים כלכליים וחשבונאיים, בין היתר לשם איתור ממוקד של כיווני ביקורת. כמו כן נבחנת יכולתן של מערכות אלה להשתלב בעבודת הביקורת באופן שיטתי ולאפשר למבקרים להתמקד בסוגיות מהותיות ובסיכונים משמעותיים יותר. לפיכך נציע מתודולוגיה סדורה ומסגרת כללית (Framework) לשימוש בכלי בינה מלאכותית בביקורת פיננסית.

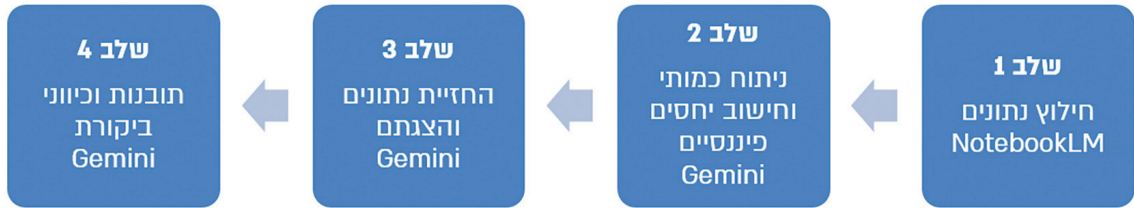
את המתודולוגיה נדגים באמצעות ניתוח מצבה הפיננסי של חברה ממשלתית ציבורית והשוואה בין תוצאות הניתוח המתבססות על מודלים של בינה מלאכותית ובין ממצאי

בעבודתם והן לצורך שימוש בכלים אלו בעבודת הביקורת עצמה. לדעת Genaro-Moya³³, יש להביא בחשבון את השינויים המבניים שמביאה עימה הבינה המלאכותית ולגבש אסטרטגיות למיצוי יתרונותיה, זאת תוך שיפור יעילות הביקורת והאפקטיביות שלה והתאמתה לאתגרים הצפויים. אסטרטגיה כזו צריכה לענות על שתי שאלות מרכזיות: אילו מערכות בינה מלאכותית יכולות לשפר את עבודת מוסד הביקורת, וכיצד יש להיערך לביקורת של מערכות אלו בשירותים הציבוריים ובגופים המבוקרים.

במאמר זה אנו מראים כיצד ניתן לרתום כלים מבוססי בינה מלאכותית לייעול ולקיצור של תהליכי הביקורת הפיננסית

33 Genaro-Moya, Lopez-Hernandez & Godz, Artificial Intelligence 33, לעיל ה"ש 30.

תרשים ב: תהליך הניתוח באמצעות כלי בינה מלאכותית



שלב 2: הניתוח הכמותי והחישובי, יחסים פיננסיים באמצעות Gemini

אחרי גיבוש בסיס הנתונים, הניתוח עובר לשלב הכמותי באמצעות הכלי. Gemini הייחודיות של Gemini בניתוח יחסים פיננסיים, כגון יחסי נזילות, רווחיות ומבנה הון, נובעת מיכולתו לא רק "לקרוא" את הנתונים, אלא גם לייצר ולהפעיל קוד תכנותי מאחורי הקלעים כדי לעבד את הנתונים הגולמיים, לחשב את המדדים הפיננסיים במדויק ולהפיק מהם תובנות באופן אוטומטי. בתהליך הזה המודל מחולל ומריץ קוד Python בסביבת עבודה מוגדרת לצורך ביצוע החישובים. פעולות אלה מאפשרות ניתוח רב-ממדי הכולל ניתוח אופקי (השוואת סעיפים לאורך זמן) וניתוח אנכי (יחסים בתוך סעיפי המאזן או ההכנסות). השימוש ב-Gemini מאפשר לאסוף נתונים העולים מהדוחות הכספיים של הגוף המבוקר, כפי שהתקבלו בשלב הקודם, ולעבד אותם לכדי חישובים מורכבים של מדדים, כגון יחס שוטר, מכפיל רווח ותשואה על ההון (Return on Equity; ROE). תוך זיהוי אוטומטי של מגמות חריגות הדורשות בדיקה מעמיקה.

שלב 3: ויזואליזציה והצגת נתונים

שלב הצגת הנתונים אינו רק עניין אסתטי, אלא חלק בלתי נפרד מהעברת המשמעות הכלכלית. Gemini משמש גם כמתרגם מידע כמותי לתובנות באמצעות יצירת גרפים וטבלאות אינטראקטיביים. היתרון המרכזי הוא היכולת לייצר קוד גרפים וטבלאות באופן התואם את צורכי הביקורת. במקום גרפים גנריים, המודל מאפשר יצירת לוחות מחוונים (Dashboards) טקסטואליים או גרפיים המדגישים כיווני ביקורת. כך ניתן להציג את תוצאות הניתוחים באופן קוהרנטי ולהנגיש נתונים מורכבים כתובנות ויזואליות ברורות המסייעות למבקר להבין את הממצאים במהירות וביעילות.

שלב 4: תובנות וכיווני ביקורת

השלב האחרון והמתקדם ביותר בתהליך הוא הפקת תובנות מהממצאים שנאספו בביקורת. בשלב זה באה לידי

דוח ביקורת בנושא שנערך בשנת 2020 ללא שימוש בכלי בינה מלאכותית. כמו כן נעריך, בסיוע כלי בינה מלאכותית ועל בסיס ניתוח נתונים מעודכן, אם יש מקום לביצוע ביקורת מעקב על תיקון ממצאיה של הביקורת משנת 2020.

מתודולוגיה לביצוע ביקורת פיננסית באמצעות כלי בינה מלאכותית של חברת גוגל

בשנים האחרונות הפעולות של הביקורת הפיננסית והניתוח הכלכלי משתנות ועוברות תהליך של אוטומציה וייעול הכולל שימוש במודלי שפה ובסוכני בינה מלאכותית (AI Agent). המתודולוגיה הנבחנת להלן מבוססת על עיבוד נתונים מספריים לצד חילוץ של מידע טקסטואלי לא-מובנה מתוך דיווחי החברה, עיבוד נתונים, חישוב יחסים פיננסיים, ניתוח כמותי של מידע גולמי לא-מובנה והסקת מסקנות. עבור תהליך זה דרושים מספר כלים של בינה מלאכותית עם חוזקות שונות, זאת כדי למנוע, ככל האפשר, הטיות אלגוריתמיות והמצאת עובדות שקריות (הזיות).

שלב 1: חילוץ נתונים באמצעות NotebookLM

השלב הראשון בכל ניתוח פיננסי הוא המעבר מדוחות גולמיים לידע בר-עיבוד. NotebookLM הוא הכלי המתאים ביותר לפעולה זו, משום שהוא מבוסס על ארכיטקטורה של יצירה מונחית אחזור (Retrieval Augmented Generation; RAG). שלא כמודלים כלליים, כלי זה פועל בסביבה סגורה ומעוגנת במקורות שהמשתמש מזין לתוכה. היתרון המובהק בשימוש ב-NotebookLM לחילוץ נתונים מדוחות כספיים טמון ביכולתו להצליב אסמכתאות ("הצלב אסמכתאות"). המערכת מאפשרת לחלץ נתונים מתוך הדוחות הכספיים, מתוך הערות הביאורים ומתוך הסקירות של דוחות הדירקטוריון, תוך שמירה על נאמנות למידע המופיע במסמכים שהוזנו ועל דיוקו. בדרך זו נשמרת רמת מהימנות גבוהה ופוחת הסיכון ל"הזיות" המודל, שכן כל פלט מגובה בציטוט מדויק מהמקורות.

שלב 2: עיבוד אנליטי וחישוב המדדים הפיננסיים

קובץ הנתונים הגולמי שהופק בשלב הקודם הוזן לכלי Gemini, תוך שימוש בהנחיה מובנית (Prompt) לביצוע ניתוחים פיננסיים לצורך בחינת מצבו הפיננסי של הגוף המבוקר. הניתוח כלל יחסים פיננסיים בולטים, כגון יחסי נזילות ויחסי כיסוי חוב. בשלב זה התוצר הוא קובץ נתונים מעובד ומאורגן המציג תחשיב מקיף של כלל היחסים הפיננסיים שנדרשו להמשך עבודת הביקורת.

שלב 3: הדמיית הנתונים (Data Visualization) והנגשתם

על בסיס המדדים שחושבו התבקש הכלי Gemini לאפיין ולייצר לוח מחוונים (Dashboard) "יעודי" (ראו דוגמה בנספח א'). לוח המחוונים נועד להנגיש את המידע הפיננסי המורכב, לזהות חריגים ומגמות רב-שנתיות ולהפוך את הנתונים המספריים לתובנות ויזואליות ברורות שניתן להסיק מהן מסקנות.

שלב 4: איתור כיווני ביקורת, השוואת ממצאים וגיבוש המלצות

בשלב המסכם נדרש המודל Gemini לנתח את הממצאים העולים מהדיווחים הפיננסיים של הגוף המבוקר לצד המדדים שחושבו בשלבים הקודמים, במטרה להצביע על כיווני ביקורת מרכזיים. לאחר מכן הונחה הכלי להשוות בין התובנות שהפיק באופן עצמאי לממצאים שהובאו בדוח הביקורת שפורסם. השוואה זו נועדה להדגים את הפוטנציאל הטכנולוגי ולהמחיש את התרומה של כלי הבינה המלאכותית לשיפור איכות הביקורת ועומקה. נסף על כך, הכלי התבקש לגבש המלצות להיתכנות ביצוע ביקורת מעקב ולהציע כיווני ביקורת עדכניים בהתבסס על נתונים של השנים האחרונות.

תוצאות

לאחר איסוף הנתונים מהדוחות הכספיים, ניתוחם ועיבודם באמצעות כלי בינה מלאכותית, נערכה השוואה בין תוצאות הניתוח ובין הליקויים והממצאים שהובאו בדוח הביקורת שפורסם. להלן לוח 2 הכולל את תוצאות ההשוואה.

בחינה של ממצאי הניתוח שבוצע באמצעות הבינה המלאכותית לעומת דוח הביקורת שפורסם מציגה תמונה ברורה של השלמה הדדית בין שתי שיטות העבודה. מתוך שמונה סוגיות מרכזיות שעלו בבחינה, נמצאה חפיפה מלאה או חלקית בחמישה ליקויים מהותיים. אלה כוללים ניתוחים כלכליים והשפעות של אירועים משמעותיים, כגון המשבר שהתרחש בשנת 2012, מבנה ההון והמינוף, התנהלות המאסדר ותת-ההשקעה ברכוש קבוע ובעלויות המימון. לצד זאת, ניכר שלכל שיטת ניתוח יש מיקוד ייחודי: ליקוי אחד שעסק בהטבות עובדים ובהתחייבויות פנסיוניות הוצג רק בניתוח של הבינה המלאכותית, ואילו שני ליקויים

ביטוי יכולת הסקת המסקנות של Gemini. לאחר עיבוד הנתונים הגולמיים מתוך NotebookLM וניתוחם הכמותי, המודל מסוגל לבצע סינתזה בין כל הממצאים. בתהליך זה עשויים להתגלות "דגלים אדומים" (Red Flags), הסיבות להם ואף כיווני ביקורת אפשריים. כאן היתרון המרכזי של Gemini הוא ביכולתו לייצר "תוכנית עבודה מבוססת סיכונים". המודל אינו מסתפק בתיאור המצב הקיים, אלא מציע כיווני ביקורת אופרטיביים, כגון סעיפים במאזן הדורשים אימות חיצוני, הערכות שווי הנראות אופטימיות מדי או כיצד שינויים מאסדרים או מקרו-כלכליים עשויים להשפיע על נתוני החברה. תהליך זה מסייע למבקר לגבש תובנות אסטרטגיות וכיווני ביקורת מבוססים וממוקדים.

השילוב בין הכלים Gemini ו-NotebookLM יוצר תשתית מקצועית איתנה המאפשרת למבקרים להעמיק בחקר הפיננסי באופן שיטתי, תוך שמירה על דיוק מרבי, יעילות ויכולת הסקה מבוססת נתונים.

השימוש בכלי בינה מלאכותית בביקורת פיננסית על חברה ממשלתית ציבורית

לצורך בחינת השימוש בכלי בינה מלאכותית בביקורת פיננסית נערכו שני ניתוחים: בניתוח הראשון נבחנו ביצועי הבינה המלאכותית אל מול דוח ביקורת פיננסית על חברה ממשלתית ציבורית שפורסם בשנת 2020 המבוסס על נתוני הדוחות הכספיים לשנים 2008 - 2019 ללא שימוש בכלי בינה מלאכותית (להלן - דוח הביקורת שפורסם); בניתוח השני נעשה שימוש בדוחות הכספיים לשנים 2020 - 2024, תוך יישום כלי בינה מלאכותית (NotebookLM ו-gemini) ונבחן האם השימוש בכלים אלו מצביע על צורך בביצוע ביקורת מעקב על הדוח משנת 2020. להלן תיאור הממצאים.

השוואה בין ביקורת פיננסית המבוססת על כלים מסורתיים לביקורת המבוססת על כלי בינה מלאכותית תחילה נותחו הדוחות הכספיים באמצעות כלי בינה מלאכותית לפי השלבים האלה:

שלב 1: חילוץ הנתונים ובניית "בסיס ידע" בר-ניתוח

בשלב זה הוזנו למערכת NotebookLM הדוחות הכספיים המלאים של החברה (להלן - הגוף המבוקר) לשנים 2008 - 2025, כפי שפורסמו בקובצי PDF באתר המרשתת של הגוף המבוקר. הזנת המידע הזה ייצרה עבור הכלי "בסיס ידע" מובנה וייעודי. מתוך בסיס ידע זה חולצו אוטומטית הנתונים הפיננסיים המרכזיים: סעיפי מאזן, דוח רווח והפסד ודוח תזרים מזומנים לצד מידע איכותני רלוונטי ומבואר. הנתונים שחולצו אורגנו במבנה אחיד והומרו לגיליון אלקטרוני (Excel) ששימש בסיס להמשך הניתוח.

העסקיים, התכנוניים והאסטרטגיים של הגוף המבוקר. השילוב בין השיטות מאפשר להרחיב את היקף הבדיקה ולהעמיק את איכות הממצאים. בעוד שהניתוח הממוחשב נשען ברובו על נתונים היסטוריים, עבודת צוות הביקורת מביאה לידי ביטוי שיקולים מקצועיים הבוחנים תוכניות עבודה עתידיות ומאפשרים לזהות סיכונים אסטרטגיים שטרם נראו בדוחות הכספיים.

הנוגעים להיעדרה של כרית ביטחון ניהולית ולסיכונים עתידיים אותרו רק על ידי צוות הביקורת.

השוואה זו מעלה תובנות מרכזיות בנוגע לשתי שיטות העבודה ולאופן שבו הן משלימות זו את זו. הבינה המלאכותית מצטיינת בזיהוי אנומליות מקרו-כלכליות מתוך נתונים, כגון הירידה החדה ברווחיות ב-2012, ואילו צוות הביקורת מתמחה בבחינה מעמיקה של ההקשרים

לוח 2: הליקויים והממצאים שהובאו בדוח הביקורת שפורסם והשוואתם לליקויים ולממצאים שנמצאו בניתוח של מודל הבינה המלאכותית

הליקוי	ממצאי הניתוח של הבינה המלאכותית	ממצאי דוח הביקורת	האם הממצאים קיימים בשני הניתוחים?
המשבר (2012) והשפעתו	אנומליה ברווחיות: ירידה חדה ברווחיות בשנת 2012 עקב משבר בגוף המבוקר	החזרי תעריף: הירידה בחוב אחרי שנת 2012 נבעה בעיקר מהחזרים רטרואקטיביים לחברה מהמאסדר	כן נקודות מבט משלימות: הבינה מלאכותית זיהתה את הירידה החדה ברווחיות, ובדוח הביקורת הוסבר התיקון המלאכותי שנעשה
מבנה ההון, המינוף ודירוג האשראי	פער מבנה ההון: מינוף שיא בשנת 2013, יחס הון עצמי למאזן נמוך, מימון צמיחה מבוסס חוב	דירוג אשראי עצמי וסיכון: סיכון נתפס כגבוה בהשוואה לחברות דומות באירופה, דירוג עצמי נמוך מאוד המעיד על סיכון גבוה ללא גיבוי ממשלתי	כן המינוף הגבוה שהבינה מלאכותית זיהתה הוא אחת הסיבות לדירוג האשראי העצמי הנמוך שתואר בדוח הביקורת
התנהלות המאסדר	חשבונות פיקוח נדחים: המאסדר נמנע מהעלאת תעריפים ומאפשר רישום של "נכסים רגולטוריים" במקום מזומן	משך חיים ממוצע נורמטיבי: המאסדר קובע משך חיים ממוצע קצר מידי לפריקט כדי למנוע העלאת תעריף, מה שמגדיל את סיכון המחזור של הגוף המבוקר	כן נמצא שמדיניות המאסדר משפיעה באופן מהותי על האיתנות הפיננסית של הגוף המבוקר כדי לשמור על תעריף נמוך לציבור
פיתוח - השקעה ברכוש קבוע	מחסור בהיקף השקעה: יחס הון נמוך לחברה עם "צורכי השקעות אדירים" ושירות חוב ש"אוכל" את היכולת להשקיע	תת-השקעה חמורה: ירידת החוב על חשבון פיתוח תוכנית ההשקעות הנוכחית אינה מפצה על חוסרים מהעבר	חפיפה חלקית הבינה המלאכותית מצביעה על הצורך בהשקעה, ובדוח הביקורת מודגש שהקטנת החוב פוגעת בפיתוח

המשך - לוח 2: הליקויים והממצאים שהובאו בדוח הביקורת שפורסם והשוואתם לליקויים ולממצאים שנמצאו בניתוח של מודל הבינה המלאכותית

עלויות המימון וניהול החוב ההיסטורי	מודל כיסוי החוב: מוזכר כ"שירות חוב כבד" בעקבות גיוס אג"ח מסיבי	ריביות חריגות: 12% מהאג"ח נושאות ריבית גבוהה מ-7% ומחציתן יותר מ-5%, כפול מריבית הגיוס השולית	חפיפה חלקית הבינה המלאכותית זיהתה בעיה בכושר ההחזר וביכולת ניהול החוב, ובדוח הביקורת מוכמתת באופן ספציפי בעיית הריביות הישנות והיקרות
הטבות העובדים וההתחייבויות הפנסיוניות	חוסר הלימה מול החוק: תרבות ארגונית וכוח של הוועד מכבידים על ההון העצמי של הגוף המבוקר	לא הוזכר בדוח	מצא ייחודי בניתוח של הבינה המלאכותית
בקרת סיכונים (כרית ביטחון)	לא זוהו ממצאים	כרית ביטחון חסרה: התכנון הפיננסי אינו משקף יחס נדרש של 1.2	מצא ייחודי בדוח הביקורת
הסיכונים העתידיים	לא זוהו ממצאים	יעדי שנת 2030: דרושה השקעה גדולה מאוד כדי לעמוד ביעדים לשנת 2030, מדובר בסיכון פיננסי מהותי להמשך פעילות החברה	מצא ייחודי בדוח הביקורת

המלצות הבינה המלאכותית לביצוע ביקורת מעקב

על בסיס ניתוח הדוחות הכספיים לשנים 2020 - 2025 נבחנו באמצעות כלי בינה מלאכותית (NotebookLM ו-gemini) האפשרות לביצוע ביקורת מעקב. תהליך הבדיקה כלל שני שלבים עיקריים: תחילה בוצע עיבוד הנתונים מהדוחות הכספיים לשנים אלה, תוך שימוש באותם כלים ומתודולוגיה שיושמו בבדיקה הראשונית; בשלב השני הוזן דוח הביקורת לכלי הבינה המלאכותית, ובהנחיה מובנית התבקש הכלי לבחון את מידת תיקונם של הליקויים. על סמך ממצאי הבדיקה התבקש הכלי לנסח המלצה עבור השאלה: האם קיימת הצדקה לביצוע ביקורת מעקב בנושא. להלן לוח 3 הכולל את תוצאות הבדיקה המשווה.

עם זאת, יש להכיר גם במגבלות הבינה המלאכותית: יכולתה לזהות את כל הליקויים או להבין את מכלול ההיבטים העסקיים אינה מלאה, והיא תלויה לחלוטין באיכות הנתונים המוזנים; המודל אינו מחליף את שיקול הדעת המקצועי האנושי ואינו יכול להעריך מידע שאינו מגולם בנתונים. כך לצוות הביקורת יש יתרון בגישה למידע איכותני ופנים-ארגוני, כגון ראיונות עם מנהלים, פרוטוקולים של ישיבות דירקטוריון ותכתובות עם המאסדר, תכנים שאינם נכללים בדוחות הכספיים ומתוכם אפשר לפענח את כוונות המאסדר או את הדינמיקה המתרחשת בתוך הגוף המבוקר; אומנם ניתוח באמצעות כלי בינה מלאכותית מאפשר לזהות שינויים במדדים ותופעות חריגות, אך הירידה לפרטי פרטים ואיתור הליקויים דורשים גישה למסדי נתונים משפטיים, לנתונים פרטניים ולעיתים גם לנתונים רגישים שאינם נכללים בקובצי הנתונים המאוחדים שסופקו למודל.

לוח 3: ממצאי הבינה המלאכותית בעניין הליקויים בדוח
שפורסם ומידת תיקונם בשנים 2021 - 2024

הליקוי	המצב בביקורת הקודמת, עד שנת 2020	המצב בביקורת המעקב, כפי שחישוב מודל הבינה המלאכותית, 2021 - 2024	מידת תיקון הליקוי
המנוף הפיננסי	חוב פיננסי גבוה, יחס הון עצמי נמוך	ירידה עקבית במינוף ושיפור יחס הון עצמי למאזן	תוקן במידה רבה
עלות החוב	עלות ריבית אפקטיבית גבוהה בהשוואה לחברות דומות באירופה	הוצאות המימון נותרו גבוהות, פרמיית הסיכון של החברה לא הצטמצמה במידה רבה	תוקן במידה מעטה
הסיבות לשיפור במצב הפיננסי	חשש שהיציבות נשענת על תקבולים חד-פעמיים	ההסתמכות על מקורות חיצוניים לכיסוי פערים תזרימיים נמשכת	לא תוקן או בטיפול
חשבונות פיקוח נדחים	חוסר שקיפות ופער בין רווח חשבונאי לתזרים מזומנים	יתרות "נכסי הפיקוח" בדוח 2024 עדיין גבוהות מאוד, המנגנון נותר מורכב	לא תוקן

לפי מודל הבינה המלאכותית, ההצדקה העיקרית לקיום ביקורת המעקב נעוצה באשליית היציבות שיוצרת מכירת הנכסים של הגוף המבוקר. במילים אחרות, זהו שיפור פיננסי זמני, וברגע שתסתיים מכירת הנכסים הגוף המבוקר עלול לעמוד מול "שוקת שבורה", שכן הוצאות המימון שלו נותרו כבדות ותלותו החיצונית לא פחתה. העובדה שהגוף המבוקר מצליח להציג רווחיות חשבונאית חיובית בזמן שתזרים המזומנים מפעילותו השוטפת אינו מכסה את צרכי השקעותיו היא נורת אזהרה המחייבת ביקורת מעקב.

כמו כן, לפי המודל, סוגיית "חשבונות הפיקוח הנדחים"³⁴ היא מהותית; כל עוד הגוף המבוקר נשען על "נכסים

מלוח 3 עולה שמתוך ארבעה ליקויים שנבדקו, ליקוי אחד תוקן במידה רבה, והיתר טופלו במידה מעטה או שכלל לא טופלו.

מודל הבינה המלאכותית ממליץ על ביצוע ביקורת מעקב, משום שאומנם ניכר שיפור גדול במדדי המינוף וההון העצמי של הגוף המבוקר, אך ניתוח מעמיק מגלה שהיציבות הפיננסית הנוכחית נשענת בעיקר על פתרונות זמניים וחד-פעמיים ולא על שיפור מבני ואיתנות תפעולית אמיתית. יתרה מזו, מרבית הליקויים המהותיים נותרו בעינם, ובעיקר אלו המשפיעים ישירות על תזרים המזומנים השוטף ועל שירות החוב לטווח ארוך.

34 נכסים או התחייבויות פיננסיים הנוצרים כאשר חברה הפועלת תחת אסדרה תעריפית רושמת הכנסות או הוצאות בתקופה נוכחית, והן יועברו ללקוחות או שייגבו מהם בתעריפים עתידיים, בהתאם להחלטת המאסדר.

רגולטוריים³⁵, שטרם תורגמו למזומן בקופה כדי להציג איתנות פיננסית, קיים נתק מסוכן בין הדוחות החשבונאיים למציאות תזרים המזומנים של הגוף המבוקר. לפיכך, מטרתה של ביקורת המעקב לוודא שההמלצות, ובראש ובראשונה נטרול ההשפעה ממכירת הנכסים מהרווח התפעולי ויצירת מנגנון תעריף בהיר יותר מול המאסדר, אכן מיושמות, וכי הגוף המבוקר מפתח יכולת אמיתית ועצמאית לעמוד בחובותיו מתוך פעילותו השוטפת ואינו מסתמך על פתרונות נקודתיים בלבד.

סיכום ומסקנות

לשימוש בבינה מלאכותית בביקורת פיננסית יתרונות רבים: הוא מאפשר אוטומציה של משימות שוטפות, תורם לייעול תהליכי העבודה ומאפשר עיבוד של כמויות גדולות של נתונים. יכולות אלה תומכות בזיהוי דפוסים, באיתור חריגות ובהפקת תובנות, ובכך מסייעות ביצירת תחזיות מדויקות יותר ובקבלת החלטות מושכלת יותר. נוסף על כך, הבינה המלאכותית יכולה לתרום לפיתוח מוצרים ושירותים חדשים שלא היו אפשריים בעבר, ופעולותיה עשויות להפחית עלויות ולצמצם את הפעולות האנושיות שמצריכות השקעה של זמן ומשאבים.

במאמר הנוכחי השתמשנו בכלי בינה מלאכותית לניתוח ולעיבוד של דוחות כספיים של חברה ממשלתית ציבורית. ממצאי הבדיקה מצביעים על כך ששילוב כלי בינה מלאכותית בניתוח דוחות כספיים של גוף מבוקר מספק תמונה רחבה יותר על אודות מצבו הפיננסי. עם זאת, אין לראות בכלים אלו תחליף לעבודת הביקורת המסורתית, הואיל והבינה המלאכותית אינה מחליפה את שיקול הדעת האנושי מקצועי אלא רק משלימה אותו.

בהתאם לכך, אנו ממליצים על שימוש בכלי בינה מלאכותית, הן ככלי משלים בביצוע ביקורת פיננסית הן ככלי תומך בקבלת החלטות. זאת ועוד, ניתוח בסיסי של דוחות כספיים

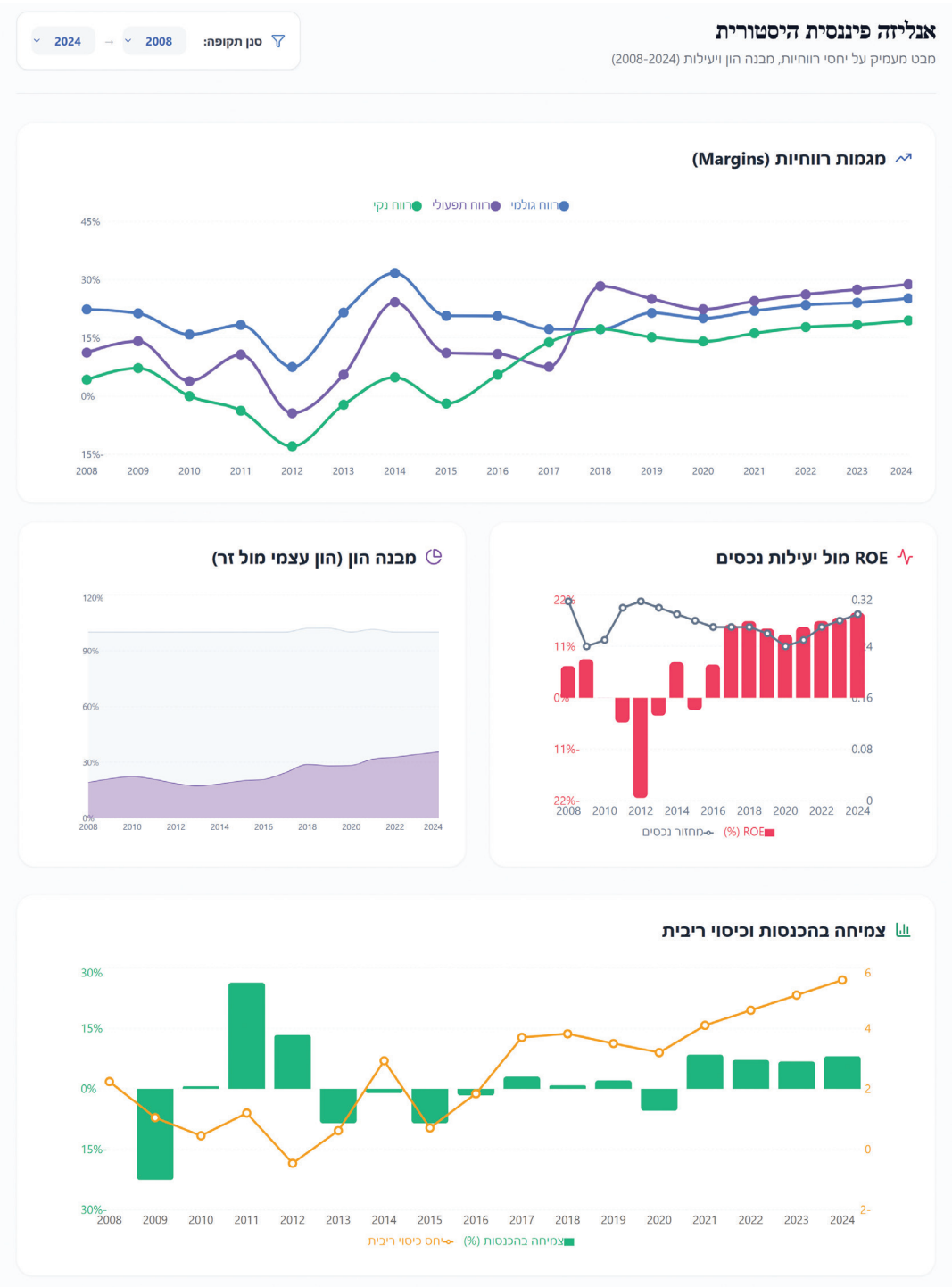
בעזרת כלי בינה מלאכותית יכול להצביע על כיווני בדיקה ועל ליקויים אפשריים ולשמש אינדיקציה למצבו העסקי של הגוף המבוקר. באמצעות הממצאים שהמודל מפיק אפשר לענות על שאלות בסיסיות כגון האם כדאי לבצע את הביקורת בכלל? אילו כיווני ביקורת כדאי לבדוק? מהם הסיכונים העיקריים שאיתם מתמודד הגוף המבוקר?

בביקורת פיננסית על חברה ממשלתית מומלץ, בשלב הראשון, להעלות את הדוחות הכספיים של החברה לכלי NotebookLM ולחלץ את הנתונים הנחוצים לתוך גיליונות אלקטרוניים. לאחר מכן על צוות הביקורת לבצע בקרה ולוודא שהנתונים שנאספו נכונים ומדויקים. בשלב הבא מועברים הנתונים שחולצו לכלי Gemini כדי לחשב יחסים פיננסיים וכלכליים, לייצר ויזואליזציה שלהם ולזהות מגמות הנובעות מהם. גם בשלב זה נדרש צוות הביקורת לבצע בקרה על החישובים, כדי לוודא את תקינותם. בהתבסס על חישובים אלה, ניתן להיעזר ב-Gemini לגיבוש כיווני ביקורת אפשריים. בד בבד מומלץ שצוות הביקורת יערוך סיעור מוחות עצמאי, יבחן את כיווני הביקורת העולים מהנתונים ויחבר בין הממצאים להקשרים העסקיים, התכנוניים והאסטרטגיים של הגוף המבוקר.

לסיכום, מומלץ להיעזר בכלי בינה מלאכותית לצורך איסוף נתונים, ניתוחם, עיבודם והצגתם באופן חזותי, שעה שצוות הביקורת ממשיך למלא תפקיד מרכזי בחשיבה, בפרשנות ובגיבוש כיווני ביקורת וליקויים, תוך חיבור בין הניתוחים הטכניים לתוצאות העסקיות והכלליות של החברה. עם זאת, יש לזכור שלצד היתרונות הרבים של השימוש בכלי בינה מלאכותית בביקורת פיננסית קיימים סיכונים. על כן ביקורת המדינה, המחויבת לשמור על אובייקטיביות, על מקצועיות ועל סטנדרטים גבוהים, נדרשת להשתמש בכלים אלו בשום שכל, תוך הקפדה על מניעת זליגה של מידע, שמירה על הפרטיות ובחינה מדוקדקת וביקורתית של תוצרי הבינה המלאכותית.

35 לדוגמה, כספים שהחברה צפויה לקבל בעתיד באמצעות התעריפים, למשל בעקבות מכירה צפויה של נכסים, תגמול בגין התייעלות שהובטח לה על ידי המאסדר וכדומה.

נסכח א' - לוח מחוונים של הנתונים





ביקורת סייבר מונחית מדדים

גיל גזית

מר גיל גזית הוא מנהל ביקורת באגף לביקורת אבטחת המידע והגנת הסייבר בחטיבה לביקורת משרדי הממשלה ומוסדות השלטון שבמשרד מבקר המדינה ונציב תלונות הציבור.

המאמר בקצרה

- מאמר זה מציג גישה בביקורת סייבר המיועדת להפיק ממצאים מדויקים ואובייקטיביים ככל האפשר על מצבם של גופים מבוקרים בהיבטי הגנה בסייבר - ביקורת מונחית מדדים.
- גישה זו נועדה בין השאר לשפר את יכולת איתור הפערים המרכזיים ולתמוך בצורה משופרת בתהליכי קבלת החלטות בגופים בנוגע לתיעדוף הטיפול בפערים ולתקצובו.
- המאמר מתבסס בין השאר על תקן NIST SP 800-55v1.
- המאמר מגדיר את נושא המדידות בביקורת סייבר, מציג ברמת-על תהליך של הגדרת מדדים, נותן דוגמאות קונקרטיות ומספק הצצה לעתיד התחום.

מבוא | מדידות בביקורת סייבר | ערכי סף ועידן ה-AI | סיכום

מבוא

בגוף מבוקר כוחה יפה גם לגופים אלה, מתוך שתי נקודות מבט: בחינת פעולתם במרחב הדיגיטלי ככל גוף מבוקר; ובחינת פעולתם בתחום הובלת ההגנה בסייבר בהתאם לייעודם ולמשימותיהם. מאמר זה לא ירחיב בנקודת המבט של ביקורת בנושאי האסדרה של גופים מסוג זה מעבר לציון כמה נושאים.

היבטים נבחנים בביקורת המדינה

כאמור, ביקורת המדינה פועלת לאיתור ליקויים ופערים ולמצאת בעיות השורש והחסמים המרכזיים לשם טיוב פעולת הגופים המבוקרים וקידום של מילוי ייעודם ושל התועלת לציבור.

כשיטה, מוגדרים שלושה סוגים של היבטים הנבחנים בביקורת המדינה:

1. היבטי ציות - Compliance Audit
2. היבטי ביצועים - Performance Audit
3. היבטים פיננסיים - Financial Audit (לא יידונו במסגרת מאמר זה)

ביקורת בהיבטי ציות מתמקדת בבחינה של הסוגיה הנבדקת **למול הנורמות** החלות עליה. בסוג זה של ביקורת בתחום ההגנה בסייבר נבדק אם פעולות הגוף המבוקר בתחום זה עומדות בנורמות החלות עליו בתחום זה, כגון תקנות הגנת הפרטיות, הנחיות מערך הסייבר הלאומי ו-ISO 27001.

ביקורת בהיבטי ביצועים מתמקדת בבחינת עמידתו של הגוף המבוקר בעקרונות של חיסכון (לא תידון במסגרת מאמר זה), **יעילות - Efficiency** (הפקת המרב מהמשאבים הזמינים) ו**מועילות - Effectiveness** (עמידה ביעדים ובמטרות שהוצבו) בסוגיה הנבדקת - במקרה זה פעילות הגוף המבוקר בתחום ההגנה בסייבר.

ביקורת יכולה לשלב כמה היבטים בתוך נושא הביקורת.

נושא המאמר

מאמר זה מציג גישה בביצוע של ביקורת בסייבר. מטרת הגישה היא להפיק ממצאים מדויקים ואובייקטיביים ככל האפשר על מצבם של הגופים המבוקרים בהיבטים הנבחנים, וכך לשפר את יכולת איתור הפערים המרכזיים ולתמוך במידה משופרת בתהליכי קבלת החלטות בנוגע לתעודין הטיפול בפערים ולתקצובו. מטרה זו תושג באמצעות **מידות כמותיות** של היבטים הנבחנים בנושא הביקורת.

בשנים האחרונות, מתוך הבנת המשמעות האסטרטגית של הנושא, שם מבקר המדינה מר מתניהו אנגלמן דגש על ביקורות בתחום ההגנה בסייבר. ביקורות סייבר מתבצעות הן כביקורות ייעודיות ואם העניין מתבקש - כפרקים בתוך ביקורות בנושאים אחרים.

יעד מרכזי של ביקורת המדינה הוא טיוב פעולת הגופים המבוקרים וקידום של מילוי ייעודם ושל התועלת לציבור¹. לשם כך ביקורת המדינה פועלת לאיתור ליקויים ופערים ולמצאת בעיות השורש והחסמים המרכזיים.

היבטי הסייבר בייעודם ובפעולתם של הגופים המבוקרים

הגופים המבוקרים פועלים וממלאים את ייעודם גם **במרחב הדיגיטלי**, היבטי מינהלה, היבטי שירות לציבור, בהיבטים אופרטיביים (כגון OT) ובהיבטים נוספים.

אסטרטגיית הסייבר הלאומית שפרסם מערך הסייבר הלאומי בפברואר 2025 מגדירה את **החזון הלאומי** בתחום זה כ"מרחב דיגיטלי **אמין, זמין ובטוח** המאפשר את היתרונות של עולם משגשג ומתקדם". המסמך מגדיר את שלושת המרכיבים העיקריים הכלולים בו:

אמינות - היכולת לסמוך על אותנטיות המידע, על התוכן ועל הזהויות הפועלות במרחב.

זמינות - המבטיחה נגישות רציפה למרחב המקוון ולשירותים הדיגיטליים לכל שכבות האוכלוסייה והארגונים.

בטיחות וביטחון (Safety and Security) - המתייחסים להגנה מקיפה מפני איומי סייבר על תפקוד המשק, ובכלל זה על הפעולות המתבצעות במרחב הדיגיטלי, על הנכסים הדיגיטליים, ועל פרטיות המשתמשים וביטחונם, מהאזרח ועד הרמה הלאומית.

מכאן שמרחב דיגיטלי אמין, זמין ובטוח **מאפשר** את פעולת הגופים ואת מילוי ייעודם. לאור זאת, בהתאם לאסטרטגיית הסייבר הלאומית, אפשר להגדיר את **המטרה** של פעולת ההגנה בסייבר בגוף מבוקר כהשגת **אמינות, זמינות וביטחון** בנגזרת של המרחב הדיגיטלי שבה פועל הגוף, **בהתאמה** לייעודו של הגוף ולמשימותיו, **לנוכח איומי סייבר**.

סוג ייחודי של גופים המבוקרים בנושאי סייבר הם גופים אסדרתיים מדינתיים שתפקידם להנחות בתחום אבטחת המידע וההגנה בסייבר בהתאם לחוק ולהחלטות הממשלה. ההגדרה לעיל של מטרת פעולת ההגנה בסייבר

1 משרד מבקר המדינה ונציב תלונות הציבור **מדריך לביקורת המדינה**, <https://www.mevaker.gov.il/state-audit/roles-work/state-audit-guide>, תפיסת ביקורת המדינה, חזון מבקר המדינה.

היומי ברמת חומרה (Severity) גבוהה (הקריטריון) וברמת חומרה בינונית ומטה.

המונח **מדד** במאמר זה יוגדר **כמהותו (תיאורו) של ההיבט הנבחן** בתהליך המדידה.

בהגדרת מדדים לביקורת סייבר יש להתייחס לשני מאפיינים:

1. המרכיבים הנבדקים - מושאי המדידה
2. סוגי המדדים

מרכיבים נבדקים בבחינת הגנה של גוף

מסמך NIST CSF 2.0 הוא מסגרת (Framework) של קווים מנחים בנושא ניהול וצמצום של סיכוני סייבר בארגון. המסגרת מארגנת את תפוקות אבטחת המידע בארגון, ברמה הגבוהה ביותר, באמצעות שש פונקציות (ראו תרשים א):

1. **משילות ארגונית - Govern**: ביסוס אסטרטגיית ניהול סיכוני סייבר בארגון
 2. **זיהוי - Identify**: הבנה של הארגון את סיכוני הסייבר שלו
 3. **הגנה - Protect**: יישום הגנות על נכסי המידע לצורך צמצום הסבירות (Likelihood) להתרחשות אירועי סייבר וצמצום ההשלכות (Impacts) שלהם
 4. **גילוי - Detect**: זיהוי וניתוח של נסיונות תקיפה (מוצלחים או כושלים) ושל חולשות סייבר בארגון
 5. **תגובה - Respond**: טיפול בנסיונות תקיפה ובחולשות סייבר שזוהו
 6. **התאוששות - Recover**: חזרה לפעילות וצמצום ההשפעות של אירוע סייבר שהתרחש
- המסגרת מפרטת לכל אחת מהפונקציות קטגוריות (Categories) ותת-קטגוריות (Sub-Categories) של תפוקות אבטחת מידע בארגון². המסגרת גם מכנה לנומרות משלימות כגון NIST SP 800-53 (בקורות אבטחת מידע) ו- NIST SP 800-30 (מדריך לביצוע הערכת סיכוני סייבר).

המאמר מסתמך, בין השאר, על תקן NIST SP 800-55v1², שפורסם בדצמבר 2024 ועוסק במדידת אבטחת מידע. התקן נועד להיות מדריך לפיתוח ובחירה של מדדי אבטחת מידע ברמות הארגון, המשימה והמערכת כדי לזהות את מידת הצלחתם של מסמכי מדיניות, נהלים ובקורות הקיימים בארגון בתחום.

מדידות בביקורת סייבר

בדיקות הערכה (Assessment)

חלק מההיבטים הנבחרים בביקורת מבוססים, לפעמים בהכרח, על הערכות איכותניות או הערכות כמו-כמותיות. למשל רמות נזק מפגיעה בסייבר בנכס מידע, כפי שהן מוגדרות על ידי הגוף, מבוטאות בדרך כלל בסולם של שלושה או חמישה ערכים (1 - נמוכה, 2 - בינונית, 3 - גבוהה, וכד'). בחינת יישום של גוף את ההנחיה (החוקית, האסדרתית או הרגולטורית) למיפוי נכסי מידע יכולה להיות מבוטאת גם היא בסולם של שלושה או חמישה ערכים (1 - במידה מלאה, 2 - במידה חלקית, 3 - בכלל לא וכד'). הערכות כאלה הן סובייקטיביות ופרשניות ודורשות הסכמה בין הגורמים המעורבים.

מדידה (Measurement) והערכה כמותית (Quantitative Assessment) בביקורת סייבר

מדידה בביקורת סייבר היא תהליך ההפקה של ערכים כמותיים המאפשרים להאיר - להביא למודעות - **היבט** מסוים בפעילות ההגנה בסייבר בגוף המבוקר, באופן מדויק ואובייקטיבי ככל האפשר (מובן שהיבט מסוים שנבחן הוא בהגדרתו **חלק** מתמונת מצב כוללת. יתרה מזאת, גם **אוסף** של היבטים נבחרים הוא עדיין רק חלק מתמונת מצב כוללת, בהתאמה לביקורת בכללותה, הבוחנת **נושא** מסוים).

הערכה כמותית בוחנת תוצאות של מדידה מול **קריטריון** או **מטרה** שנקבעו, והיא **סוג של מדידה**. אפשר למדוד, למשל, את מספר ההתרעות המתקבלות ב-SOC³ במשך יום. אפשר לבצע הערכה כמותית של שיעור ההתרעות

2 NIST National Institute of Standards and Technology - הוא מוסד ממשלתי במשרד הכלכלה של ארצות הברית המפרסם תקנים שארגונים מקצועיים ברחבי העולם מקבלים על עצמם כאמות מידה מקצועיות לתפקוד. מערך הסייבר הלאומי מקבל את תקינת NIST כתקינה מומלצת במסמכים מקצועיים שהוא מפרסם, ובכלל זה בתורת ההגנה 2.0 ובחקיקה כגון חוק התמודדות עם תקיפות סייבר חמורות במגזר השירותים הדיגיטליים ושירותי האחסון (הוראת שעה - חרבות ברזל), התשפ"ד-2023.

3 Security Operations Center; SOC

4 תורת ההגנה 2.0 של מערך הסייבר מציינת בנספח ג' את NIST CSF כבסיס להגדרת מבנה בנק הבקורות. סייבר ישראל, מערך הסייבר הלאומי **תורת ההגנה לנהל את הסיכון: המדריך היישומי (השלם) להגנת הסייבר של הארגון (תוה"ג 2.0)** (2021), https://www.gov.il/he/pages/cyber_security_methodology_2



על פי NIST CSF 2.0, תרשים 2.

מסגרת זו ונורמות מקובלות אחרות יכולות לשמש בבחירת המרכיב הנבדק שעבורו יוגדר מדד. אפשר לבחור, למשל, במרכיב ID.RA-05 מתוך NIST CSF 2.0, שמשמעותו **שימוש של הארגון בנתוני תרחישי איום, חולשות, סבירות למימוש איומים והשלכות כדי להבין את הסיכונים ולתעדף את הצמצום שלהם** או במרכיב PR.PS-05 מתוך NIST CSF 2.0, שמשמעותו **מניעת התקנה והרצה של תוכנה לא מורשית**.

דוגמאות למרכיבים נבדקים בגופים אסדרתיים מדינתיים בתחום ההגנה בסייבר

בגופים אסדרתיים מדינתיים בתחום ההגנה בסייבר ייגזר המרכיב הנבדק מייעודם ומהגדרת משימותיהם. בגוף אסדרתי שעם משימותיו נמנות הנחיה של גופים בתחום ההגנה בסייבר ובקרה על יישום ההנחיות אפשר, למשל, לבחון את הבקורות שעשה, לבחון את המעקב שעשה אחר תיקון ליקויים שעלו בבקורות קודמות ועוד.

סוגי המדדים

תקן NIST SP 800-55v1 מבחין בין ארבעה סוגים של מדדים:

1. מדדי יישום - Implementation
2. מדדי מועילות - Effectiveness

3. מדדי יעילות - Efficiency

4. מדדי השפעה - Impact

מדדי יישום - Implementation

מדדי יישום מראים את ההתקדמות ביישום של בקורות אבטחת מידע. ביישום של אמצעי למניעת הרצה והתקנה של תוכנה לא מורשית (אנטי-וירוס, XDR, EDR וכו') אפשר להגדיר, לדוגמה, מדד של מספר התחנות בארגון שעליהן מותקנת הגרסה האחרונה של תוכנת ההגנה הארגונית ביחס למספר התחנות הכולל בארגון שבהן נדרשת התקנה של אמצעי זה.

מדדי מועילות - Effectiveness

מועילות מוגדרת כעמידה ביעדים ובמטרות שהוצבו. מדדי מועילות מעריכים את **שיעור ההצלחה** של תהליכי היישום ושל בקורות אבטחת מידע במושגים של **עמידה ביעדים ובמטרות** שהוגדרו להם.

במידה מסוג זה יש לשים לב להגדרת היעדים והמטרות בארגון. גוף יכול להגדיר לעצמו **יעד** של התקנת אמצעי עדכני למניעת הרצה והתקנה של תוכנה לא מורשית (אנטי-וירוס, XDR, EDR וכו') על כלל המחשבים בגוף. עמידה של הגוף ביעד זה אין משמעותה בהכרח שהושגה **המטרה** שלשמה מותקן האמצעי (כגון פעולה בטוחה של הגוף במרחב הדיגיטלי למול איום סייבר של הרצה והתקנה של תוכנה לא מורשית וחשיפה של מידע חסוי). מתוך ההבנה כי הרצה והתקנה של תוכנה לא מורשית מסכנות את חסיון המידע, אפשר בהקשר זה להגדיר מדד של השגת **המטרה** כמספר התרעות האמת התקבלו לאורך זמן ב-SIEM⁵ על התקנה או פעולה של תוכנה לא מורשית בתחנות הארגון.

מדדי יעילות - Efficiency

יעילות, כאמור, מוגדרת כהפקת המרב מהמשאבים הזמינים. מדדי יעילות יבחנו היבטים כגון מהירות הפקה של תוצרים שימושיים מיישום של בקורות אבטחת מידע וכמויות של תוצרים שימושיים מיישום בקורות אבטחת מידע ביחס למשאבים המושקעים ביישומם.

דוגמה למדד יעילות יכולה להיות ממוצע הזמן להשלמת טיפול בהתרעות בסייבר (MTTR - Mean Time To Resolve/Repair). מדד זה בוחן את משך הזמן של סך הפעילויות מרגע התרחשות אירוע, דרך הופעה של התרעות רלוונטיות ב-SOC, ניתוח ההתרעות, תיעוד ופעולות נוספות עד להשלמת הטיפול באירוע ככל שאכן היה כזה.

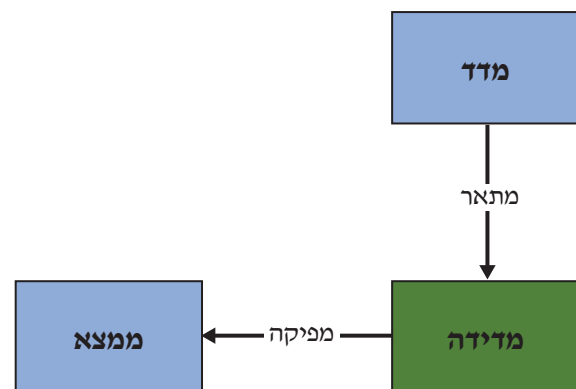
מדדי השפעה - Impact

מדדי השפעה מבטאים את ההשלכה של פערי הגנה בסייבר על מטרות הארגון, על משימותיו ועל היעדים שהוא מגדיר לעצמו. לדוגמה, בבית חולים שנמצאו בו פערי הגנה בסייבר מול תרחיש של כופרה, כגון שיעור נמוך של תחנות בארגון שבהן מותקנת התוכנה העדכנית ביותר למניעת הפעלה והתקנה של תוכנה לא מורשית, אפשר לבחון את מספר ימי ההשבתה של בתי חולים שבהם התרחשו מקרים דומים ולהעריך את העלות של מספר ימי ההשבתה הממוצע על בית החולים הנבדק.

ממצא

ממצא של מדידה במאמר זה יוגדר כערך כמותי אובייקטיבי המתקבל מתהליך המדידה⁶ (ראו תרשים ב).

תרשים ב: יחסי גומלין בין המונחים **מדידה**, **מדד** ו**ממצא**



מידת החשיבות של ממצאים

בתהליך ביצוע הביקורת, וקל וחומר בסיכומה, יש צורך להבין את מידת החשיבות של הממצאים המסתמנים או המדדים המגובשים בתוך הנושא הכולל, זאת כהשלמה להבנת הממצאים עצמם. ביקורת תשאף לאתר את המדדים שהם מרכזיים לנושא ושבעניינם המרחק בין הרצוי למצוי של הממצאים המסתמנים הוא הגדול ביותר, ולהשקיע בהם.

גופים מבוקרים שמקבלים ממצאי ביקורת יכולים להפיק תועלת מהצגת מידת חשיבותם של הממצאים בתוך נושא הביקורת כתכונים התומכים בתהליך קבלת ההחלטות בעניין תיעודף הטיפול בפערים והקצאת המשאבים הנדרשים.

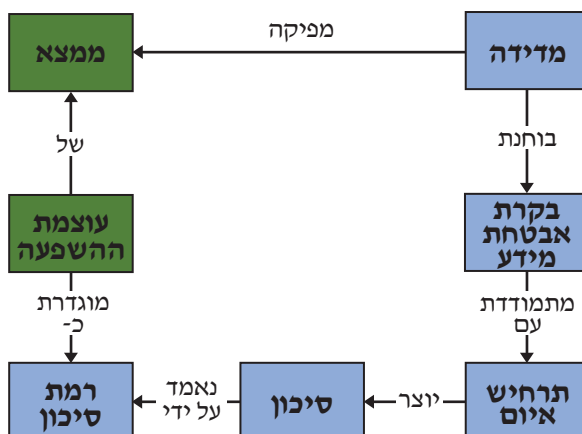
מידת החשיבות של ממצאים בתוך נושא הביקורת תוגדר **כעוצמת ההשפעה** של הממצאים על פעולת הגוף ועל מילוי ייעודו.

חישוב עוצמת ההשפעה של ממצאים

תקן NIST SP 800-55v1 מציג נושא של תיעודף מדדים⁷. הדרך העיקרית המוצעת בתקן היא תיעודף של מדדים באמצעות מודל סבירות והשפעה או, במלים אחרות, רמת הסיכון המוערכת לתרחיש האיום שבקרת אבטחת המידע הנמדדת נועדה להתמודד איתו. בשיטה זו נדרש להעריך (או לקבל מתוך ניתוח מאושר של תרחישי איום) את רמת הסבירות להתממשות תרחיש האיום המדובר ואת השפעתו - רמת הנזק שתיגרם לגוף בקרות התרחיש - ולשקלל אותן לרמת סיכון⁸ (ראו תרשים ג, ד).

דרך זו נמצאת בהלימה עם הגדרת המטרה של פעולות ההגנה בסייבר בגוף מבוקר - כהשגת **אמינות, זמינות וביטחון** בנגזרת של המרחב הדיגיטלי שבה פועל הגוף, בהתאמה לייעודו של הגוף ולמשימותיו, **לנוכח איומי סייבר**.

תרשים ג: אופן חישוב עוצמת ההשפעה של ממצא



6 ייתכנו גם ממצאים של בדיקות המבוססות על הערכה. במקרה זה ממצא יהיה איכותני או כמו-כמותי.

7 סעיף 4.3 Prioritizing Measures. National Institute of Standards and Technology (NIST), Measurement Guide for Information Security: 4.3 Volume 1 - Identifying and Selecting Measures (2024), <https://csrc.nist.gov/pubs/sp/800/55/v1/ipd>

8 ראו בעניין זה תקן NIST SP 800-30r1, לוח NIST, Guide for Conducting Risk Assessments, 2012; סייבר ישראל, מערך הסייבר הלאומי, תורת ההגנה לנהל את הסיכון, לעיל ה"ש 4, בעמ' 36.

תרשים ד - דוגמה לחישוב רמות סיכון בתלות ברמת סבירות להתרחשות איום וברמת הנזק

רמת סבירות	1	2	3	4	5
5 - גבוהה מאוד	1	2	3	4	5
4 - גבוהה	1	2	3	3	4
3 - בינונית	1	2	2	2	3
2 - נמוכה	1	1	1	2	2
1 - נמוכה מאוד	1	1	1	2	2

תרשים נמדד

5 - גבוהה מאוד 4 - גבוהה 3 - בינונית 2 - נמוכה 1 - נמוכה מאוד

רמת נזק

התרשים מבוסס על NIST SP 800-30r1, לוח I-2.

במקרים כאלה רצוי לנתח ולזהות בקורות אבטחת מידע המושפעות מהממצאים העקיפים ולעשות מדידות השלמה של יישומן על ידי הגוף המבוקר. אם יימצא מתאם בין הממצאים העקיפים לממצאי מדידות ההשלמה, ואם אפשר יהיה לבסס כי הממצאים העקיפים הם גורם משמעותי לממצאים שעלו במדידות ההשלמה, אפשר יהיה להתייחס לממצאי המדידות המשלימות כבעלי זיקה ישירה, להסיק את רמת הסיכון כאמור לעיל ולשייך אותה לממצאים העקיפים (ראו תרשים ה).

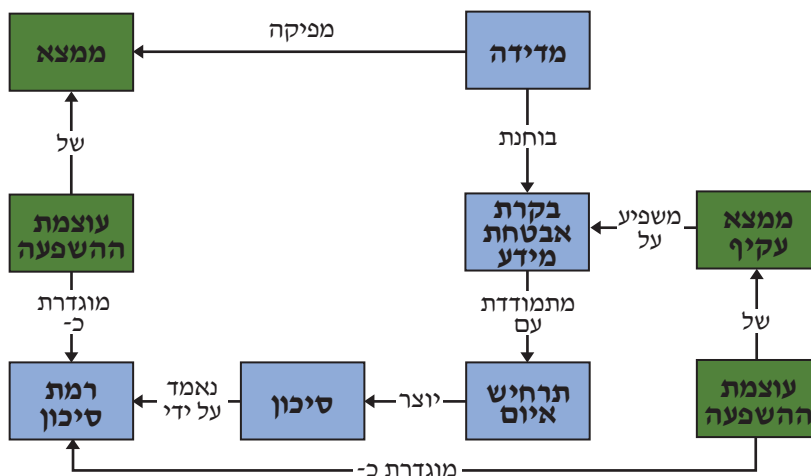
נניח לדוגמה שנעשתה בדיקה של יישום ההנחיה למיפוי נכסי מידע בארגון ולסיווגם (סעיפים 6.1.3, A.8.1.1 ו- A.8.2.1 בתקן ISO27001) ונניח שהבדיקה העלתה כי נעשה מיפוי חלקי של נכסי מידע. במקרה זה אפשר לנתח ולזהות נכסי מידע שאינם מופיעים במיפוי הנכסים - למשל מחשבים ניידים של מנהלי מערכות - ולעשות בדיקת השלמה

למשל, למדידת מספר התרעות האמת שהתקבלו ב-SIEM לאורך זמן על התקנה או פעולה של תוכנה לא מורשית בתחנות הארגון יש זיקה ישירה לתרחיש איום של הפעלת קוד מפגע וחשיפה של מידע חסוי (וגם לתרחישים נוספים). בדוגמה זו נניח שהמדידה העלתה ממצא של מספר גבוה יחסית של התרעות, ונניח שרמת הסיכון שנותחה בגוף לתרחיש המדובר היא גבוהה, עוצמת ההשפעה של הממצא תהיה גבוהה - בהתאם לרמת הסיכון.

מדידת עוצמת השפעה של ממצאים בזיקה עקיפה לתרחישי איום

ממצאים של בדיקות מסוימות, כגון ממצאים של בדיקות ציות, אינם נמצאים בזיקה ישירה לתרחישי איום. ממצאים כאלה כשלעצמם לא מעידים בהכרח על פער או היעדר פער בהתמודדות עם תרחיש איום כלשהו.

תרשים ה: אופן מדידת עוצמת ההשפעה של ממצא עקיף



חלקית, ובעקבות זאת לא יהיה אפשר לעשות מדידות מיטביות, למשל במצבים של חסר במערכות הגנה בסייבר, לוגים שאינם מכילים מידע רלוונטי ולוגים שאינם נשמרים לאורך זמן.

ערכי סף ועידן ה-AI

מהפכת ה-AI משפיעה השפעה של ממש גם על עולם ההגנה בסייבר. בין ההשפעות אפשר לציין הורדה של רף הכניסה של פושעי סייבר שבעזרת כלי AI יכולים ללא ידע מעמיק לבצע תקיפות, ומכאן לעלייה במספר התוקפים; הנגשת יכולות תקיפה מתקדמות; עליה בקצב ובמהירות של תקיפות; ושיפור ביעילות של תקיפות.⁹

התפתחות זו מעלה היבט נוסף הקשור במדדים - ערכי סף. כאשר עושים מדידות כמותיות אפשר להגדיר ערכי סף המבחינים בין מצב תקין ובין פער. בלי להידרש לאופן הגדרת ערכי סף, אפשר לומר שיכולות התקיפה המבוססות על AI מקצינות את ערכי הסף. למשל, אם קצב התקיפות עולה ויעילות התקיפות עולה, מהירות התגובה להתערות סייבר צריכה לעלות בהתאם, כך שאפשר יהיה לסכל את התקיפה בזמן.

ההתפתחות לעיל מעלה היבטים נוספים, כמו מידת היכולת של בני אדם להתמודד עם קצב התקיפות, שימוש בכלי AI להגנה כדי להתאים את קצב המענה, ומדידת היעילות והמועילות של מערכות הגנה מבוססות AI. התמודדות עם התפתחויות אלה הן חלק מאתגרי העתיד של משרד מבקר המדינה.

סיכום

לביקורת המדינה, ובפרט בהיבטי סייבר, כמה יתרונות. יתרון אחד הוא עצמאות הביקורת, האובייקטיביות שלה ואי-התלות שלה¹⁰. יתרון נוסף הוא הגישה של משרד מבקר המדינה לאלפי גופים מבוקרים ברמה של כלל המדינה, המאפשרת הצלבה של נתונים, ניתוח של קשרי גומלין והצגת תמונת מצב ומגמות רוחביות.

לשימוש בגישה של ביקורת מונחת מדדים יש יתרונות הן למשרד מבקר המדינה והן לגופים המבוקרים, והיא מוצגת ככלי אפשרי בביצוע ביקורת סייבר.

של יישום בקורות אבטחת מידע על אותם נכסים. נניח שבדיקות ההשלמה מעלות את הממצא שעל מחשבים אלה לא יושמה הבקרה של התקנת תוכנות למניעת קוד מפגע, ונניח שתרחיש האיום המשווין לבקרה זו הוא ברמת סיכון גבוהה. אפשר להניח כי היעדרם של המחשבים הניידים הללו ממפיו הנכסים תרם במידה ניכרת לאי-יישום הנחיה מתאימה ולא-ביצוע מעקב אחר יישום הבקרה של תוכנת ההגנה כפי שמבוצע במסגרת תהליכי ניהול אבטחת המידע בגוף. לכן עוצמת ההשפעה של הממצא העקיף בעניין מפיו נכסי מידע היא גבוהה בהתאם לרמת הסיכון המשווית לתרחיש האיום הרלוונטי.

אם הביקורת עשתה בדיקות השלמה ולא העלתה ממצאים המושפעים מהממצא העקיף אי אפשר לקבוע את רמת החשיבות של הממצאים העקיפים כעוצמת ההשפעה שלהם על פעולת הגוף ועל מילוי ייעודו בזמן ביצוע הביקורת. עם זאת, בדוגמה המדוברת אין להמעיט בחשיבותם של ממצאי בדיקות ציות כשלעצמם (בדגש על חוקים ורגולציות), ובפרט בהקשר של עמידת הגוף במטרות פעולות ההגנה בסייבר לאורך זמן, של שימור ההגנה בסייבר בגוף ושל ההתקדמות בנושא בעת כניסתם של בעלי תפקידים חדשים לתפקיד, בדגש על מנהלים ועל נושאים שיתכן שלא נבדקו באותה ביקורת.

היבטים של בשלות הגוף המבוקר בתחום ההגנה בסייבר

לרמת הבשלות של גוף בהיבטי הגנה בסייבר השפעה על המדידות הניתנות לביצוע.

יישום בקורות אבטחת מידע

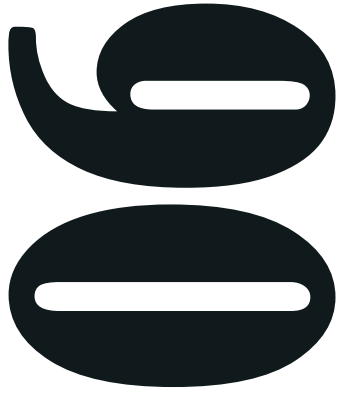
ככלל, אי אפשר לבדוק היבטי מועילות ויעילות של בקורות אבטחת מידע שגוף אינו מיישם כלל. מתוך כך - ניתן לבצע כשלב ראשון בביקורת בדיקות ציות שמטרתן לזהות בקורות רלוונטיות שאינן מיושמות כלל, וכאמור לעיל, לבצע בדיקות משלימות הבוחנות את ההשפעה של הבקורות החסרות על פעולת הגוף ומילוי ייעודו.

יכולות מדידה

מדידות אבטחת מידע מתבססות, בין השאר, על יכולות של גוף מבוקר לשמור נתוני אבטחת מידע בעזרת אמצעים כגון קבצי לוג של מערכות אבטחת מידע (Firewall, IDS, XDR) וכו'. בגופים שרמת הבשלות שלהם בתחום ההגנה בסייבר נמוכה ייתכן שלא ייושמו אמצעים כאלה או שיושמו במידה

9 מאיה נחום שחל "אנחנו רגע לפני מלחמת הסייבר הראשונה שבה לא תהיה ירייה אחת" כלכליסט (19.11.2025) calcalist.co.il/conferences/article/ryhca0fgbx

10 משרד מבקר המדינה ונציב תלונות הציבור **מדריך לביקורת המדינה**, לעיל ה"ש 1, תפיסת ביקורת המדינה, עקרונות הליבה <https://www.mevaker.gov.il/state-audit/roles-work/state-audit-guide>



איומי סייבר על מערכות ביטחוניות בלתי מאוישות בעידן הבינה המלאכותית הצעה למסגרת ביקורת מעודכנת להערכה ולבקרה

איתן דהן, משה מורנו, אבישי יהושע ואורן בר

תא"ל (במיל') איתן דהן הוא ראש החטיבה לביקורת מערכת הביטחון במשרד מבקר המדינה ונציב תלונות הציבור.
סא"ל (במיל') משה מורנו הוא סגן מנהל האגף לביקורת צה"ל, כשירות ומוכנות בחטיבה לביקורת מערכת הביטחון שבמשרד מבקר המדינה ונציב תלונות הציבור.

ד"ר אבישי יהושע הוא מנהל ביקורת באגף לביקורת העורף האזרחי בחירום, גבולות ואיו"ש בחטיבה לביקורת מערכת הביטחון שבמשרד מבקר המדינה ונציב תלונות הציבור.

אל"ם (במיל') אורן בר הוא מנהל ביקורת באגף לביקורת משרד הביטחון ומשאבי מערכת הביטחון בחטיבה לביקורת מערכת הביטחון שבמשרד מבקר המדינה ונציב תלונות הציבור.

המאמר בקצרה

- מערכות בינה מלאכותית (להלן - AI) משולבות כיום במערכות ביטחוניות בלתי מאוישות, לרבות בתהליכי מודיעין, שליטה ובקרה וזיהוי מטרות, ובכך הן מרחיבות את משטח התקיפה של המערכת מעבר לתשתיות הקלאסיות (חומרה ותוכנה) - אל הנתונים, המודלים ושרשרת האספקה של רכיבי AI.
- מאמר זה עוסק בשאלה כיצד יש לעדכן את עבודת הביקורת בתחום הגנת הסייבר לאור העובדה שרכיב הליבה של המערכות הביטחוניות הבלתי מאוישות - מנגנון קבלת ההחלטות - נשען על מודלים ונתונים דינמיים ומבוסס על AI.
- המאמר ממפה וקטורי תקיפה מרכזיים ומראה כיצד כשלים אלה עלולים להתגלגל מכשל טכני לכשל מבצעי ולכשל אחריותיות.
- בהמשך המאמר מוצעת מסגרת של הבטחת מוכנות לביקורת (Audit-Ready Assurance) המותאמת לביקורת המדינה, המבוססת על עקרונות הבטחת משימה וחוסן תפעולי: קריטריונים, מדדי ביצוע וסיכון, סט ראיות מזערי, לצד תוכנית בדיקות ודגימות.
- תרומת המאמר היא המעבר מביקורת "ציות לבקורת" לביקורת על "כולת הוכחה" לשם "הבטחת משימה" - שבה ממצא, מסקנה והמלצה נשענים על ראיות ניתנות לשחזור לאורך מחזור חיי המערכת.
- מילות מפתח: AI (בינה מלאכותית), מערכות בלתי מאוישות (UxV), הבטחת משימה (Mission Assurance), סייבר, ביקורת המדינה, סחיפת נתונים, אחריותיות (Accountability).

מבוא | מפת האיומים ומשטחי התקיפה במערכות ביטחוניות בלתי מאוישות | השפעת איומי סייבר ייחודיים בעידן ה-AI | מודל הבטחת משימה וחוסן מבצעי במערכות ביטחוניות בלתי מאוישות בעידן ה-AI | מסגרת ביקורת להערכה ולבקרה המוכוונת להבטחת משימה (Audit-Ready Assurance Framework) בעידן ה-AI | יישום המסגרת: מתודולוגיית ביקורת עדכנית והמלצות לשיפור מתמשך | סיכום

ביטחוניות בלתי מאוישות

מערכות ביטחוניות בלתי מאוישות בעת הנוכחית אינן עוד יחידות קצה מבודדות, אלא "מערכות מורכבות של מערכות" (System of Systems) המקושרות ברשתות נתונים הדוקות ופועלות בסביבה מבוזרת המרחיבה את היקף משטחי התקיפה. ניתוח מעמיק של משטחי התקיפה מעלה כי כל שלב במחזור החיים של המערכת - החל בפיתוח, דרך ההטמעה וכלה בעדכוני הקושחה⁴ (Firmware) - הוא נקודת תורפה אסטרטגית.

כדי לאפשר ביקורת אפקטיבית בתחום הסייבר בעידן ה-AI מוצעת מפת איומים ומשטחי תקיפה פוטנציאליים בראייה רב-שכבתית: (א) שכבת תשתיות (מערכות מחשוב כדוגמת שרתים) ותקשורת נתונים; (ב) שכבת נתונים, חיישנים ומערכות ניווט; (ג) שכבת מודל והסקה - הרצת אלגוריתמי AI ברכיבי קצה; (ד) שכבת שרשרת אספקה דיגיטלית; (ה) שכבת תפעול (מערכות שליטה קרקעיות) וממשקי אדם-מכונה. לכל שכבה מוגדרים וקטורי תקיפה אופייניים, מנגנון הפגיעה והשלכה משימתית, ובצמוד להם - סט ראיות מזערי המאפשר למבקר להוכיח שליטה בסיכון (כפי שיפורט בהמשך).

להלן הרחבה על מיפוי האיומים ומשטחי התקיפה הפוטנציאליים:

1. **שכבת תשתיות ותקשורת נתונים:** ערוצי התקשורת (Data Links) בין הכלי ובין המפעיל האנושי או תחנת הבקרה הם יעד ראשון במעלה לשיבוש. תוקף מתוחכם עלול לבצע "הזרקת פקודות" (Command Injection) שמוסוות כתקשורת לגיטימית, המנצלת את הגישה ליכולות ה-AI ככלי תומך החלטה בידי המפעיל האנושי כדי להשתלט על הכלי מרחוק.
2. **שכבת נתונים, חיישנים ומערכות ניווט:**⁵ מערכות ביטחוניות בלתי מאוישות חשופות להפרעות ניווט על

בעשורים האחרונים הצטיידו גופי הביטחון למיניהם במערכות ביטחוניות בלתי מאוישות מתקדמות לשימוש מבצעי (מערכות UXV). מערכות ביטחוניות בלתי מאוישות הן מערכות אוטונומיות הפועלות ללא נוכחות אנושית ישירה, והן נשענות במידה הולכת וגדלה על מודלים של למידת מכונה (ML)² לצורך ניווט, זיהוי מטרות וקבלת החלטות בזמן אמת. מערכות כאלה משמשות למטרות שונות, ובהן איסוף מודיעין, תקיפה מהאוויר, שליטה ובקרה. בעידן הנוכחי מערכות בלתי מאוישות הן רכיב ביטחוני מרכזי שבו מצטלבים חדשנות טכנולוגית, רציפות תפקודית, צמצום הסיכון לחיי אדם ואחריותיות ארגונית. לשימוש במערכות אלה יתרונות רבים לצד סיכונים מתפתחים. מאמר זה עוסק בשאלה כיצד יש לעדכן את עבודת הביקורת בתחום הגנת הסייבר לאור העובדה שרכיב הליבה של מערכת ביטחוניות בלתי מאוישת - מנגנון קבלת ההחלטות - נשען על מודלים ונתונים דינמיים. המאמר משלב סקירת מסגרות תקינה וניהול סיכונים AI, מיפוי וקטורי תקיפה מרכזיים בסביבת מערכות ביטחוניות בלתי מאוישות וגזירת קריטריונים ראייתיים לביקורת. הטענה העיקרית היא כי ביקורת בנושא הגנה על מערכות ביטחוניות בלתי מאוישות מפני איומי סייבר המתמקדת בבקורות הגנת סייבר מסורתיות בלבד - סודיות (Confidentiality), אמינות/שלמות (Integrity), זמינות (Availability)³ - אינה מספקת, משום שהיא אינה בוחנת שליטה ראייתית בנתונים, במודלי AI ובשרשרת האספקה הדיגיטלית של קבלני המשנה המתבססים על מרכיבי AI, ולכן אינה מאפשרת להעריך יכולת עמידה במשימה ואחריותיות.

המאמר ינתח את מפת האיומים החדשה, יבחן את סיכונים ה-AI הייחודיים המשיקים לסיכונים סייבר "קלאסי", יציג מודל להבטחת משימה (Mission Assurance) ויציע מסגרת של הבטחת מוכנות לביקורת (Audit-Ready Assurance Framework) המותאמת לאתגרים הדינמיים של שדה הקרב העתידי.

1 "X" - Aerial, Ground, Surface UXV - Underwater Unmanned "X" Vehicle;

2 Machine Learning; ML

3 המשולש של אבטחת המידע (CIA Triad) הוא המודל הבסיסי והחשוב ביותר בעולם הגנת הסייבר. שלושת המושגים הללו - סודיות, אמינות/שלמות, זמינות - מייצגים את היעדים העיקריים שכל מערכת אבטחה שואפת להשיג כדי להגן על מידע ומערכות. Jennifer Cawthra et al., Data Integrity: Detecting and Responding to Ransomware and other Destructive Events (NIST Special Publication 1800-26A, December 2020), <https://doi.org/10.6028/NIST.SP.1800-26>.

4 קושחה היא תוכנית בסיסית הצרובה על רכיב החומרה ומשמשת כעין "מוח" שלו. תפקידה הוא לנהל את פעולתו, לתרגם פקודות ולגשר בין החומרה לרכיבי התוכנה הגבוהים יותר.

5 Mark L. Psiaki & Todd E. Humphreys, GNSS Spoofing and Detection, 104(6) Proceedings of the IEEE 1258 (2016).

השפעת איומי סייבר ייחודיים בעידן ה-AI

איומי סייבר בעידן ה-AI נבדלים מאיומי סייבר קלאסיים, שכן הם מכוונים ללוגיקה של הלמידה וההסקה, ולא רק לניצול חולשות קוד. שלוש משפחות מרכזיות של איומי סייבר יש בעידן הזה: (1) "הרעלת נתונים" בשלב האימון והעדכון של מודל ה-AI (הן בשלב הפיתוח והן בתפעול השוטף של המערכת), המשנה את גבולות ההחלטה של מודל ה-AI (יפורט בהמשך); (2) מניפולציית קלט למודלים של AI (תקיפות קלט לעומתיות בזמן אמת - המטות סיווג או החלטה באמצעות שינוי זעיר בקלט); (3) סחיפת מודלים ונתונים¹¹ על ידי שינוי הסביבה המבצעית לנתונים שהמודל לא אומן עליהם. מבחינת הביקורת המשמעות היא שהערכת עמידות של מודל ה-AI מחייבת ראיות לשליטה בנתונים ולתיעוד תהליך שינוי מודל, לרבות בדיקות חוסן מתועדות.

להלן פירוט של סיכוני הסייבר העיקריים העלולים להשפיע על הבטחת המשימה הביטחונית בעידן ה-AI במערכות ביטחוניות בלתי מאוישות המשתמשות ב-AI:

1. "הרעלת נתונים" (Poisoning Data)¹² בשלב הפיתוח או העדכון של מודל ה-AI: זהו שיבוש מכוון של שלב האימון והעדכון של המודל. לדוגמה, תוקף מחדיר נתונים מוטעים למאגר המידע של המערכת הביטחונית הבלתי מאוישת במהלך פיתוח או בזמן עדכונה במשך מחזור חייו של המאגר. בעקבות זאת המודל מפתח "עיוורון" מובנה או הטיות ספציפיות המופעלות רק בתנאים מסוימים. בביקורת יש לבחון את "ניקיון" הנתונים המשמשים לאימון מודל ה-AI במערכות ביטחוניות בלתי מאוישות.

כי מערכות ניווט גלובליות בחלל (GNSS⁶) מסוג חסימה (Jamming), וחשבוש וחמור מכך - הטעיה (Spoofing). היכרות של תוקף עם חולשות מודלי AI המוטמעים ברכיבי קצה המשרתים את המערכות הבלתי מאוישות (Edge AI) עלולה לאפשר לו לשדר קלט שגוי שמודל ה-AI אינו רגיש אליו ולפגוע בהבטחת המשימה. הטעיית ניווט יכולה לשבש, למשל, את יכולת ה-Edge AI של הכלי לחשב מסלול בטוח, מה שמוביל להתרסקות או לסטייה של המשימה לכיוון אזורים מאוכלסים.

3. שכבת מודל והסקה - הרצת אלגוריתמי AI ברכיבי קצה (Edge): בניגוד למערכות ענן, רכיבי קצה במערכות ביטחוניות בלתי מאוישות נדרשים להריץ אלגוריתמי AI בזמן אמת על גבי המערכת האוטונומית עצמה. תוקפים יכולים לנצל תופעה זו לביצוע תקיפות "מניעת שירות" אלגוריתמיות⁸, המעמיסות על המעבד עד להשבתת יכולת קבלת ההחלטות של הכלי.

4. שכבת שרשרת האספקה הדיגיטלית: זוהי אולי נקודת התורפה הקריטית ביותר. התלות בספקים חיצוניים ובעדכונים שוטפים של תוכנה וקושחה לשם עדכון יכולות ה-AI מייצרת סיכון של החדרת "דלת אחורית" (Backdoor) כבר בשלב הפיתוח⁹.

5. שכבת תפעול (מערכות שליטה קרקעיות, GCS)¹⁰ וממשקי אדם-מכונה: מערכת השליטה הקרקעית היא הצומת המרכזי של המידע. חדירת נזקה למערכת זו עלולה לא רק להשבית את המערכת, אלא גם לאפשר מניפולציה של הנתונים המוצגים למפעיל האנושי ולהוביל לקבלת החלטות מבצעיות שגויות על בסיס מידע כוזב.

6 Global Navigation Satellite-based Systems

7 מודל AI המוטמע ברכיבי המותקן בכלי האוטונומי, בשונה מיכולת הנמצאת בענן או בתחנת שליטה קרקעית.

8 Jelena Mirkovic & Peter Reiher, A Taxonomy of DDoS Attack and DDoS Defense Mechanisms, 34(2) ACM SIGCOMM Computer Communication Rev. 39 (2004)

9 Tianyu Gu, Brendan Dolan-Gavitt, & Siddharth Garg, BadNets: Identifying Vulnerabilities in the Machine Learning Model Supply Chain, ARXIV (August 2017), arXiv:1708.06733 [cs.CR]; Bolun Wang et al., Neural Cleanse: Identifying and Mitigating Backdoor Attacks in Neural Networks, 2019 IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 2019, pp. 707-723, doi: 10.1109/SP.2019.00031.

10 Ground Control Systems

11 תופעה שבה המציאות והנתונים בעולם משתנים עם הזמן, מה שגורם למודל ה-AI - שהוכשר על נתונים ישנים או קבועים - להיות פחות מדויק ופחות רלוונטי למשתמשים.

12 Battista Biggio, Blaine Nelson, & Pavel Laskov, Poisoning Attacks Against Support Vector Machines, ICML'12: Proceedings of the 29th International Conference on Machine Learning 1467 (2012); Jacob Steinhardt, Pang Wei Koh, & Percy Liang, Certified Defenses for Data Poisoning Attacks, Proceedings of the 31st International Conference on Neural Information Processing Systems 3520 (2017).

להלן שני מקרי בוחן להדגמת סיכון סייבר המבוסס על מניפולציית קלט למודלים של AI:

1. אמצעי נשק קרקעי בלתי מאויש (כדוגמת מערכת "רואה-יורה" שבשימוש כוחות הביטחון) המוצב להגנה על מתקן ממשלתי רגיש, המשתמש ב-AI לזיהוי מטרות, חווה מתקפת "מניפולציית קלט למודל":

א. התוקף מציב דפוס חזותי על הקרקע, שאינו נראה לעין האנושית, הגורם לאלגוריתם הראייה הממוחשבת של מודל ה-AI לסווג מטרה אזורית כאובייקט עיון, או להפך.

ב. בעקבות זאת מערכת הנשק עלולה להמליץ על תקיפה שגויה שתוביל לפגיעה בחפים מפשע. כשל כזה אינו רק טכני; הוא כשל ב"מינהל התקין" ובאחריותיות של המדינה כלפי החברה.

2. **מניפולציית קלט למודלים של (Model Manipulation) AI¹³:** הטיית פלט המערכת בזמן אמת באמצעות שינויים מזעריים בקלט, שאינם נראים לעין אנושית (Adversarial Examples). לדוגמה, עיבוד מידע חזותי (תמונות, תצלומי אוויר וכיו"ב) שגוי בשלב הקלט המאפשר לסווג כלי רכב צבאיים ככלי רכב אזרחיים עלול לגרום למודל AI לסווג רכב צבאי כרכב אזרחי ובכך למנוע מהמערכת לבצע את משימתה המבצעית.

3. **סחיפת מודלים ונתונים (Data Drift):** זהו כשל תפקודי הנובע משינוי קיצוני בסביבה המבצעית שמודל ה-AI לא אומן עליה. איומי סייבר יכולים לנצל מצבי "סחיפה" כדי להוביל את המערכת הביטחונית הבלתי מאוישת לקריסה בתנאי קצה. דוגמה לכך היא הצבת חוסמי/ מטעי GNSS¹⁴ במרחב הפעולה המונעים מפלטפורמות UXV לנווט בצורה מדויקת.

תמונה 1: מערכת "רואה-יורה" להגנת הגבול



המקור: הופק באמצעות Gemini-i Chat-GPT

Aleksander Madry et al., Towards Deep Learning Models Resistant to Adversarial Attacks, 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings (2018); Francesco Croce & Matthias Hein, AutoAttack, ARXIV 2 (March 2020), arXiv:2003.01690 [cs.LG].

14 אמצעי המוצב על הקרקע ומונע או משבש קליטה של אותות ניווט לווייניים ממערכות לווייני ניווט (Global Navigation Satellites Systems).

ג. כשל זה ממחיש כיצד חולשה בלוגיקת ה-AI הופכת לכשל מבצעי קריטי, גם כאשר "חומת האש" ותשתית התקשורת נותרו חסינות לחלוטין.

מודל הבטחת משימה וחוסן מבצעי במערכות ביטחוניות בלתי מאוישות בעידן ה-AI

בהקשר של מערכות ביטחוניות בלתי מאוישות "הבטחת משימה" מורכבת מסט יכולות מדידות: המשך ביצוע משימה בתנאי שיבוש, ירידה מבוקרת למצב בטוח (fail-safe) (להלן - אל-כשל) וזמן התאוששות מוגדר.

ג. התוקף האדפטיבי (Adaptive Adversary) ינצל בדיוק את אותן חולשות אלגוריתמיות כדי לעקוף את מנגנוני ההגנה הסטטיים של המערכת.

2. תקיפת סייבר על נחיל רחפנים בניסוי סימולציה:

א. נחיל רחפנים אוטונומי להגנה על תחנת כוח, המצויד במערכת זיהוי מטרות סיגינטית¹⁵ (על פי תדר שידור) מבוססת AI, הוטעה באמצעות "משדרי הטעיה" שהותקנו על כלי רכב ידידותיים לכאורה.

ב. השינוי המזערי בתבנית השידור, שאינו נקלט בחיישן הקולט המותקן על כל אחד מהרחפנים בנחיל, גרם לאלגוריתם לסווג את הרכב "מטרה צבאית עוינת".

תמונה 2: נחיל רחפנים המגן על תחנת כוח במרחב אזרחי



המקור: הופק באמצעות Gemini-1 Chat-GPT

15 סיגינט (Signal Intelligence; SIGINT) - מודיעין אותות, ענף בעולם המודיעין העוסק באיסוף מידע על ידי יירוט ופיענוח של אותות אלקטרוניים והפקת מידע מודיעיני מתוכם (אתר המרכז למורשת המודיעין).

מילוי מלא ותקין של כלל מרכיבי המשימה בעידן השפעת ה-AI מחייב ביתר שאת מעבר מתפיסת "הגנה סייבר היקפית" ו"הגנה על המערכת" לתפיסת הבטחת משימה וחוסן מבצעי (Mission Assurance & Operational Resilience) - היכולת להשלים את המשימה גם בתנאי שיבוש או תקיפה מוצלחת. לכן הביקורת נדרשת לבדוק לא רק "אם יש בקורות", אלא אם הארגון מסוגל להוכיח - בראיות ובמדדים - עמידה ביעדי שירות והמשך תפקוד בתרחישי שיבוש, לרבות באמצעות תרגילי שרידות ותיק אירועים.

עלינו להניח כי המערכת הביטחונית הבלתי מאוישת תהיה נתונה לשיבוש או לחדירה, ולכן המיקוד בפיתוח המערכת, בהוכחת יעילותה ובתפעולה השוטף עובר למיקוד ביכולתה להשלים את המשימה חרף הפגיעה.

יש לנהל את הסיכונים למערכת הביטחונית הבלתי מאוישת לאורך כל שלבי מחזור החיים שלה:

1. **פיתוח:** הטמעת בקורות (Security by Design) ואימון אדברסרי של מודל ה-AI (אימון הכולל מימוש סיכוני סייבר אפשריים ודרך להתמודד עימם).
2. **הטמעה:** בדיקות קבלה (Acceptance Testing) לחוסן מודל ה-AI¹⁶.

3. **תפעול:** ניטור רציף (Continuous Monitoring) של החלטות מודל ה-AI במערכת הביטחונית הבלתי מאוישת.
 4. **עדכונים:** תיקוף של מודל ה-AI לאחר כל עדכון נתונים או קושחה.
- נוסף על כך, לשם הבטחת המשימה נדרש לקיים את "מעגל החוסן המבצעי" המורכב מארבעה שלבים קריטיים:
1. **מניעה:** צמצום משטחי התקיפה, אימות חתימות דיגיטליות של עדכוני קושחה והבטחת "היגיינת נתונים" בתוכם בשלב הפיתוח של מודל ה-AI.
 2. **זיהוי:** ניטור מתמיד וחכם של חריגות סטטיסטיות בהחלטות מודל ה-AI בהשוואה ללוגיקה המצופה.
 3. **תגובה:** הפעלת מנגנוני אל-כשל המבטיחים כי המערכת הביטחונית הבלתי מאוישת המבוססת על מודל AI לא תפעל באופן המסכן את הסביבה.
 4. **התאוששות:** יכולת חזרה לתפקוד, עדכון מודלי ה-AI בשטח (Over-the-air updates) ויכולת תחקור מלאה של האירוע.

תרשים א: מעגל החוסן המבצעי



המקור: הופק באמצעות Gemini-1 Chat-GPT

16 בדיקות קבלה (Acceptance Testing) לחוסן המודל הן השלב הקריטי שבו מוודאים שמודל ה-AI מדויק לא רק בשלב התכנון, אלא עמיד בפני תרחישי קיצון, מניפולציות ורעשים בעולם האמיתי, לפני הטמעתו המלאה במערכת מבצעית או עסקית.

מסגרת ביקורת להערכה ולבקרה המוכוונת להבטחת מוכנות לביקורת בעידן ה-AI (Audit-Ready Assurance- Framework)

מיפוי תקנים בין-לאומיים וישראליים לצורך ביקורת הגנה מפני איומי סייבר במערכות ביטחוניות בלתי מאוישות

התקינה לביקורת ולהערכה של בקורות בתחום. אבטחת מידע והגנת סייבר היא רחבה ומעמיקה. בעידן של מערכות ביטחוניות בלתי מאוישות מבוססות AI נדרש להרחיב את נקודת המבט כך שלצד בקורות הגנת סייבר קלאסיות יהיו גם בקורות שיבחנו גם שלמות נתונים, שינויי מודל ה-AI, חוסן והטיה וכן שרשרת אספקה של רכיבי AI.

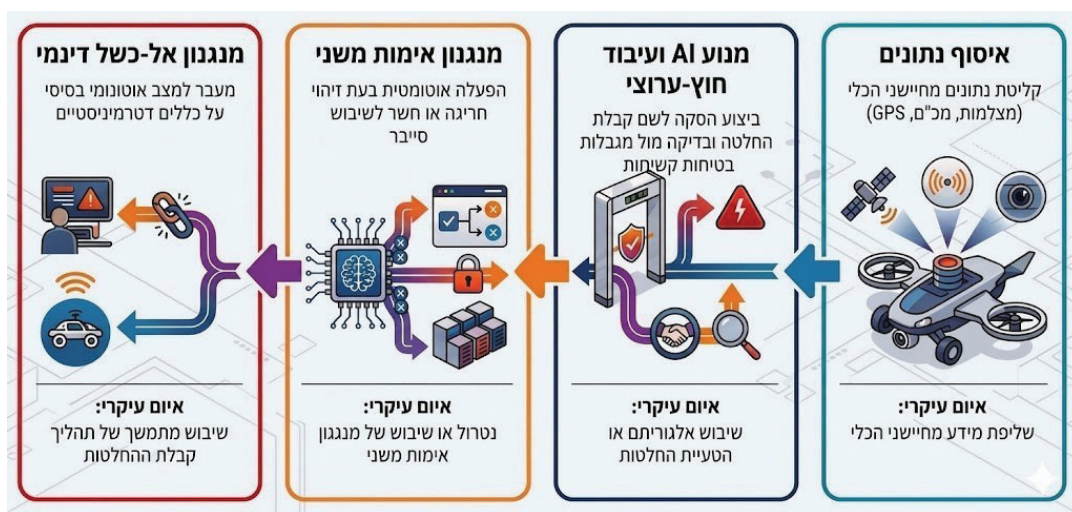
לצורך יצירת מסגרת ביקורת עדכנית מוצעת חלוקה פונקציונלית של התקנים לשלוש קבוצות, שתאפשר לבנות מסגרת Ready-Audit שאינה מחליפה תקינה קיימת, אלא מוסיפה שכבת ראיות וקריטריונים הממוקדים בנתונים ובמודל שרשרת אספקה דיגיטלית: (א) תקנים לניהול סיכונים סייבר ובקורות; (ב) תקנים ייעודיים לניהול סיכונים AI; (ג) תקנים למתודולוגיות לעבודת ביקורת.

תהליך הבטחת משימה במערכות ביטחוניות בלתי מאוישות

התהליך מתחיל בשלב איסוף הנתונים מחיישני הכלי (מצלמות, מכ"ם, GPS). הנתונים מוזרמים אל מנוע ה-AI המרכזי המבצע הסקה (Inference) לשם קבלת החלטה. לאחר קליטת הנתונים מחיישני הכלי וקבלת החלטה על ידי מודל ה-AI, מופעל מנוע ניטור חוץ-ערוצי (of-Out-Monitoring band) המשווה בין החלטת ה-AI ובין מגבלות בטיחות קשיחות. אם מזוהה חריגה (למשל: זיהוי מטרה בתוך שטח אסור) המערכת הביטחונית הבלתי מאוישת עוברת למנגנון אימות משני.

עם החשד לשיבוש סייבר מתאמת המערכת מפעילה מנגנון אל-כשל (safe-Fail) דינמי: ניתוק ה-AI, מעבר למצב אוטונומי בסיסי המבוסס על כללים דטרמיניסטיים (Hard-coded) ושידור התרעה דחופה למפעיל האנושי. רצף זה מבטיח את הרציפות התפקודית של המשימה המבצעית גם בתנאי "אש דיגיטלית", כלומר מתקפת סייבר מתקדמת על מודל ה-AI השזור במערכת.

תרשים ב: תהליך הבטחת משימה



המקור: הופק באמצעות Gemini-1 Chat-GPT

שואפת לחזק את הפיקוח על הגופים המבוקרים ועל תשתיות לאומיות קריטיות, תוך כדי שימוש בכלים מתקדמים כמו מבדקי חדירה.

(2) **מתודולוגיה לביקורת אבטחת מידע, הגנת הסייבר והגנת הפרטיות:** המסמך מציג את מטרות הביקורת (שמירה על סודיות-שלמות-זמינות ואבטחה של מידע אישי בהתאם לחוק ולתקנות) וקובע דגשים ארגוניים לעבודת צוות הביקורת (הכשרה ייעודית; תיאום עם האגף לביקורת אבטחת המידע והגנת הסייבר במשרד; והתייחסות מחמירה לסיווג וחסיון). המסמך מפרט גם את הסיכונים העיקריים ואת הבקורות הנגזרות מהם, מתאר את תהליך הביקורת משלב בחירת הנושא דרך זיהוי הגופים המבוקרים וקביעת נורמות הביקורת, ומפרט נושאי בדיקה שכיחים (ממשל סייבר, תוכניות עבודה ומשאבים, מדיניות ונהלים, ניהול סיכונים, אבטחה פיזית/לוגית, פיתוח מאובטח, המשכיות עסקית, שרשרת אספקה, זיהוי אירועים וטיפול בהם והגורם האנושי). עוד הוא מספק הנחיות לכתיבת דוח ממוקד ועדכני ואפשרות לשילוב מבדקי חדירה/חוסן והסתהיכות ביועצים. הוא גם מסביר את מפת האסדרה בישראל (מערך הסייבר, שב"כ, יח"ב/מערך הדיגיטל, יחידות מגריות והרשות להגנת הפרטיות) ואת בסיסי הסמכות והחובות בדיון, לרבות עדכונים בהגנת הפרטיות.

עדכון התקינה הקיימת

התקנים הקלאסיים נבנו לעולם של "תוכנה + רשת + שרתים". במערכות ביטחוניות בלתי מאוישות מבוססות AI יש עוד שלושה רכיבים קריטיים שהתקנים בדרך כלל לא נדרשו להם מספיק: הנתונים, מודל ה-AI והשרשרת הדיגיטלית שמספקת אותם.

לנוכח זאת מוצע העיקרון שלהלן: כשמערכת ביטחונית בלתי מאוישת נשענת על נתוני AI ועל מודל AI, הביקורת נדרשת להיעזר בכלי ניתוח נתונים שיאפשרו למבקר

בנספח למאמר רוכזו תקנים בין-לאומיים עיקריים המקובלים להתוויית המתודולוגיה לביקורות בתחום אבטחת המידע והסייבר בארגונים וכמה תקנים העוסקים במסגרת ייעודית לניהול סיכוני AI ותקינה ברמה הלאומית ובמשרד מבקר המדינה. להלן עיקרם:

1. מסגרת ייעודית לניהול סיכוני AI

א. **NIST AI RMF¹⁷ 1.0 (NIST AI 100-1) (2023):** מסגרת ניהול סיכוני AI לפי Manage/Measure/Map/Govern; מגדירה מאפייני ארון (Trustworthiness) וניהול סיכונים במודלים ובנתונים לאורך מחזור חיים.

ב. **MITRE¹⁸, Adversarial Threat Landscape for Artificial Intelligence Systems (ATLAS):** מיפוי של כל הטקטיקות והטכניקות שתוקפים עשויים להפעיל לאורך כל מחזור החיים של מערכת AI - החל בשלב איסוף הנתונים וכלה בשלב הפעולה (Inference).

ג. **ISA¹⁹/IEC 62443 (IACS²⁰/OT²¹):** סדרת תקנים לאבטחת מערכות בקרה ותעשייה או טכנולוגיות הפעלה (IACS/OT) הרלוונטיים במיוחד לפלטפורמות אוטונומיות, לרכיבי קצה ולשילוב OT-IT במערכות Physical-Cyber.

2. מתודולוגיות לעבודת ביקורת

א. **מערך הסייבר הלאומי - תורת ההגנה בסייבר לארגון (מסמך תפיסתי-יישומי):** אחריות ההנהלה, הגנה לפי פוטנציאל נזק, מניעה-זיהוי-תגובה וחזרה לשגרה, בסיס לחיבור בין הנחיה לביקורת.

ב. **משרד מבקר המדינה:**

(1) **תפיסת המשרד בנוגע לביקורת טכנולוגיות מידע והגנת סייבר:** המסמך מציג את התפיסה האסטרטגית של משרד מבקר המדינה בנוגע לביקורת טכנולוגיות מידע והגנת הסייבר. הוא מפרט את המבנה הארגוני, הכולל הקמת אגפים ייעודיים לצד הכשרת כוח אדם מקצועי בכל חטיבות הביקורת, להתמודדות עם איומים דיגיטליים. על פי התפיסה, מדיניות המשרד

Risk Management Framework; RMF 17

MITRE 18 - ארגון אמריקאי המנהל מרכזי מחקר ופיתוח במימון ממשלתי ללא כוונת רווח.

International Society of Automation; ISA 19

Industrial Automation and Control Systems; IACS 20

Operational Technology; OT 21

ב. בקרת שינוי מודל (Model Change Control): בקרה על כל עדכון של מודל ה-AI שמחויב בתהליך אישור (בדומה לשינוי בכל מערכת קריטית).

ג. בדיקות לפני הפעלה (Pre-deployment Tests): לפני שמכניסים מודל חדש יש לבדוק אותו מול "סט בדיקות קבוע" (קבוצת תרחישים) כדי לוודא שלא "נשברה" התנהגות חשובה.

ד. בדיקת יכולת חזרה אחורה (Rollback): בדיקה זו חשובה מכיוון שאם אחרי עדכון מתגלה בעיה נדרשת יכולת להחזיר גרסה קודמת במהירות. נוסף על בדיקת יכולת חזרה לאחור, יש לוודא שמתקיימים תרגולים תקופתיים בנושא זה.

3. ללא תוספות אלו לתקינה הקיימת, הארגון עלול "להחליף מוח" למערכת אוטונומית בלי להבין מה השתנה.

4. יש לשלב בתקינה הקיימת מדדי התנהגות ייעודיים ל-AI: Drift, שגיאות קריטיות ואמון (ראו פירוט להלן). בתקנים הקלאסיים מודדים זמינות, הרשאות ואירועי סייבר. במערכות ביטחוניות בלתי מאוישות אוטונומיות מבוססות AI נדרש למדוד גם "בריאות התנהגותית" - האם המודל עדיין מתאים למציאות. מומלץ להוסיף לתקינה את מדדי התנהגות ה-AI הבאים:

א. Drift (סטייה לאורך זמן): המודל עלול להידרדר בלי "להתריע", ולכן יש לבדוק אם הנתונים שהמערכת רואה היום שונים מהנתונים שעליהם אומנה.

ב. שיעור שגיאות קריטיות: לא רק "כמה טעויות", אלא "אילו טעויות".

ג. מדד אמון: נדרש לוודא שיש נוהל ברור המנחה על הפעולה הנדרשת אם המודל אינו אמין, לדוגמה הפניה לבקרה אנושית, חייושן נוסף או עצירה.

5. הסיכון במערכות ביטחוניות בלתי מאוישות הוא לא רק "פריצה", אלא גם "הידרדרות שקטה" המובילה להחלטות שגויות.

6. בתקנים רבים זמינות נתפסת כיעד IT כללי ("המערכת זמינה"). במערכת ביטחונית בלתי מאוישת אוטונומית, גם אם "השרת חי" אבל הנתונים בזמן אמת מגיעים באיחור או מגיעים באופן חלקי - המערכת למעשה אינה כשירה. לכן בתקינה הקיימת נדרש לעדכן דרישות זמינות ושרידות כלהלן:

לזהות חריגות ולהסביר את ההחלטות שהתקבלו על ידי מודל ה-AI. בהיבט זה ראוי לנסח דרישת סף: יכולת שחזור (Reproducibility) על דגימה מצומצמת ותיעוד ברור של תהליך שינוי נתוני AI/מודל AI.

על כן העדכונים המוצעים להלן הם בגדר תוספות לתקינה, ולא החלפה שלה:

1. יש להוסיף לתקינה "שכבת נתונים" מחייבת: מקור, איכות, תיוג וגרסאות (ראו פירוט להלן). תקני אבטחת מידע מסורתיים מתמקדים ב"הגנה על מידע" (הרשאות, הצפנה, גיבוי), אבל פחות ב"נכונות המידע, עקיבותו וגיבוי בעובדות מקוריות" - שאלה קריטיות למערכות ביטחוניות בלתי מאוישות מבוססות AI, שכן מודל ה-AI לומד את הנתונים ומתפקד לפיהם. אם הנתונים במערכת ביטחונית בלתי מאוישת מבוססת AI שגויים, המערכת "תבין" את העולם בצורה שגויה. לפיכך מומלץ לעדכן בתקינה את התוספות הבאות:

א. מקור - תיעוד מקור הנתונים (Data Provenance): לתעד לכל מאגר נתונים שמשמש לאימון/הפעלה מה מקורו, מהו מועד יצירת המאגר ומועדי עדכוננו, מי אסף, באילו תנאים ומה מותר/אסור לעשות בו.

ב. איכות - בחינת איכות (Data Quality): סט של בדיקות איכות, למשל אם חסרים נתונים, אם יש חריגות, כפילויות או הטיות.

ג. תיוג - בקרה על תיוג (Labeling) מבוקר: אם יש "תיוג", למשל סימון אובייקטים בתמונה, מוכרח להיות תהליך מבוקר: מי מתייג, מי מאמת, ומה שיעור השגיאות.

ד. גרסאות - וידוא ניהול גרסאות לנתונים (Dataset Registry): כל שינוי במאגר הנתונים מוכרח להירשם כגרסה: מה השתנה, מי אישר, ומה ההשפעה.

2. יש להוסיף לתקינה "שכבת מודל" מחייבת: תוכנה רגילה מתעדכנת לפי נוהל "שינוי גרסה". שינוי קטן במודל AI עלול לשנות התנהגות באופן לא אינטואיטיבי. לכן צריך תקינה ברורה לניהול מודל ה-AI - מי אחראי למודל ה-AI, איך מעדכנים אותו, ואיך חוזרים לאחור. מומלץ לעדכן את התקינה הקיימת בתוספות הבאות:

א. רישום מודלים (Model Registry): רישום של "תעודת זהות" של מודל ה-AI: שם, גרסה, תאריך, מי אחראי, באילו נתונים אומן ולמה הוא מיועד.

א. יעדי רמת שירות (SLO²²): לא רק ש"השרת עובד", אלא שהנתון מגיע בזמן, בקצב הנדרש ובאיכות הנדרשת. מודל AI שאופיין ופותח על בסיס רמת שירות מסוימת עלול להיטק מרמת שירות השונה באופן מהותי, ובעקבות זאת לפגוע בהבטחת המשימה.

ב. תרגילי ניתוק ועומס: לפחות אחת לפרק זמן יש לדמות נתק תקשורתי או עומס כדי לבדוק איך המערכת מתנהגת לאחר התקלה.

ג. Graceful Degradation: תקן המחייב התנהגות בטוחה במצב חסר, לדוגמה ירידה לפעולה שמרנית, מעבר למוד "המתנה" או דרישת אישור אנושי.

ד. להוסיף נספח זמינות בתקנים כך שיקלו עיכוב (latency), עדכניות (freshness) וחלופת נסיגה (fallback) ובחינתם במבחנים מוקדמים בשלב הוכחת המערכת.

7. יש להרחיב את התקינה בנושא שרשרת אספקה כך שתכלול "שרשרת אספקה של AI": ארגונים רבים מסתמכים כיום על ספריות תוכנה, מודלים חיצוניים, שירותי ענן או רכיבי צד ג'. בתקנים קלאסיים יש התייחסות לספקים, אבל לא תמיד ברזולוציה של "מודל/ספרייה/דאטה-סט"²³ מומלץ לעדכן את התקינה הקיימת ולהוסיף ביקורת על הנושאים הבאים:

א. "עץ מוצר" (SBOM²⁴) לרכיבים קריטיים: לדעת בדיוק מאילו רכיבים בנויה המערכת.

ב. תהליך קבלה לרכיב (Acceptance): לפני הכנסת רכיב חיצוני יש לבצע בדיקות בסיס, תיעוד, אישור.

ג. דרישות חוזיות עם הספק: זמני תיקון, חובת דיווח פגיעויות ושקיפות ברכיבים.

ד. סעיף AI Components, כגון מודלים וספריות, וכן מסמכים נדרשים רלוונטיים בנושא זה בנוהלי העבודה עם הספקים.

8. יש לעדכן את התקינה כך שתכלול דרישות ביקורת ייעודיות לשימוש ב"סוכני AI" שפותחו ואומנו על ידי ספקי מודלים חיצוניים: כאשר היכולת האלגוריתמית אינה נשלטת במלואה בידי הגוף המבוקר, יש לחייב "שכבת הבטחה ספקית" כתנאי כשירות, הכוללת לכל

הפחות: (א) תיעוד תיחום שימוש (Scope) ומגבלות ידועות של הסוכן/המודל והוראות שימוש מותר/אסור; (ב) מנגנון עדכונים מבוקר - גרסאות, תהליך שינוי, התראות על שינויי התנהגות, ויכולת חזרה לגרסה קודמת; (ג) ראיות לבדיקות לפני הטמעה ובשגרה, לרבות בנוגע לחוסן ולתרחישי קצה; (ד) דרישות חוזיות מחייבות של שקיפות מינימלית, לרבות SLA לתיקון וקביעה של תהליך דיווח אירועים, וכן זכות לקבלת ראיות ביקורת של צד ג'. כך הביקורת תוכל להעריך לא רק "אם יש בקורות", אלא אם יש יכולת הוכחה חוזית-ראייתית מול הספק, לרבות לגבי כשירות, חוסן ועדכונים של סוכני ה-AI.

9. יש לקבוע בתקינה "סט ראיות מזערי" כתנאי כשירות

Audit-Ready Baseline: תקנים רבים כיום מאפשרים "הצהרה" על עמידה בהם ללא הצגת ראיות. בעידן ה-AI, ללא ראיות (למשל גרסאות, בדיקות, מדדים), אי אפשר לדעת אם המערכת נשלטת. לכן יש לקבוע בתקינה רשימה מזערית של מסמכים או לוגים שמוכרח להתייחס בכל מערכת AI קריטית. על הרשימה לכלול, בין היתר, מפת מערכת ותלויות (Dependencies), רישום מודל, דוחות בדיקה, דוחות זמינות ותיק אירועים.

10. בעידן ה-AI חשוב מאוד להוסיף לתקינה התייחסות שתבטא ביקורת שאינה מסתפקת רק בבדיקת מסמכים, כמפורט להלן:

א. בדיקות שחזור (Reproducibility): לבחור מודל אחד ומאגר נתונים אחד ולהראות שאפשר לשחזר את התהליך בצורה מבוקרת.

ב. דגימה חובה: בכל ביקורת - דגימת מודל/דאטה-סט/רכיב ספק.

ג. היצמדות למדדים: כל בדיקה מסתיימת במדד ובתוצר ראיתי.

11. את כלל התוספות לתקינה שתוארו לעיל יש לתרגם ל"שפה ארגונית" - בעלי תפקיד, לוחות זמנים, מדדי הצלחה: כדי שהתוספות לתקינה לא יישארו תיאורטיות, כל תוספת צריכה להיות מנוסחת בדרך מובנת ברמת ההנהלה: מי אחראי (Owner), מהו התוצר (Deliverable), עד מתי מודדים הצלחה ואיך.

22 Service Level Objective - היעד הכמותי המשקף את איכות השירות של המערכת, לדוגמה אחוז מתקציב שגיאות מותר.

23 דאטה-סט (Dataset) - אוסף מאורגן של נתונים ומידע שנאספו לצורך מחקר, ניתוח או אימון או תיקוף של מודלי AI (מילון Merriam-Webster).

24 Software Bill of Materials - רשימה מפורטת של כל המרכיבים בתוכנה מסוימת: ספריות קוד פתוח, רכיבי צד ג', גרסאות ספציפיות של כל רכיב, רישיונות.

תרשים ג: מודל נמי"ף המותאם לעולם הביקורת בעידן ה-AI



המקור: הופק באמצעות Chat-GPT ו-Gemini

מהנורמה עשוי להעיד על ניסיון "הרעלת נתונים" בשלבי הלמידה.

3. ערכי שפלי (Shapley Values): כלי זה פותר את בעיית "הקופסה השחורה" של מודל ה-AI עבור המבקר. הוא מאפשר להסביר (Explainability) את הערך (המשקל) של כל אחת מהתכונות (Features) שהשפיעו על החלטה ספציפית. לדוגמה, אם הביקורת מעלה כי קלט מחיישן LiDAR²⁶ תרם 70% להחלטה ניווט שגויה, המבקר יכול להצביע על פגיעות בחיישן הספציפי או על הסתמכות-יתר מסוכנת של המודל עליו.

4. ביקורת במגבלות סיווג: אף שאין מדובר במרכיב ייחודי לביקורת בעידן ה-AI, לנוכח רגישות המערכות הביטחוניות הבלתי מאוישות מומלץ שהביקורת תתבצע במודל של "גישה מדורגת" ותשתמש במערכות ניתוח נתונים מבדדות (Air-gapped) המאפשרות למבקרים בעלי סיווג מתאים לבחון לוגיקה והתנהגות של מודל AI בלי להוציא נתונים מסווגים אל מחוץ לארגון.

5. לצורך הבטחת האחידות מומלץ כי הגוף המבקר יציג מדדי חסינות רציפים. להלן המלצות מעשיות

מתודולוגיית ביקורת מבוססת נתונים במערכות ביטחוניות בלתי מאוישות משולבות AI

ביקורת על מערכות ביטחוניות בלתי מאוישות ששולבו בהן יכולות AI אינה יכולה להסתמך על ראיות מסורתיות בלבד. על כן מוצע לאמץ מתודולוגיית Data Science המבוססת על מודלים ועל כלי מדידה שעשויים לתמוך במסגרת הביקורת כגון:

1. מודל נמי"ף (DIKA)²⁵: לעומת המאה העשרים, שבה היה מצופה שהמבקר יבחן כיצד נתונים גולמיים (Data) הופכים למידע (Information), לידיע (Knowledge) ולתובנות מבצעיות (Wisdom), ויזהה נקודות כשל בעיבוד המידע, בעידן ה-AI נדרש לאמץ מודל DIKA, שהוא וריאציה של DIKW המכוונת לפעולה (Action). מודל זה יכול לשמש מסגרת להבנת התהליך שבו המערכות "לומדות" ומקבלות החלטות לפעולה. להלן פירוט של שלבי המעבר מנתונים לפעולה, שהוא לב ליבה של מערכת ביטחונית בלתי מאוישת משולבת AI:

א. נתונים (Data): הקלט הגולמי (Data Raw) שהמערכת מקבלת; למשל קובצי תמונות, רצפי מספרים מחיישנים או טקסטים לא מעובדים.

ב. מידע (Information): שלב העיבוד שבו ה-AI מחלץ מאפיינים (Features); לדוגמה זיהוי "כתם אדום" בתמונה או זיהוי מילים ספציפיות במשפט.

ג. ידע (Knowledge): שלב הלמידה שבו ה-AI מזהה תבניות (Patterns) וקשרים; למשל המערכת מבינה ש"כתם אדום" במקום מסוים ובצורה מסוימת הוא "תמרור עצור".

ד. פעולה (Action): השלב הקריטי ב-AI יישומי שבו הידע מתורגם להחלטה אופרטיבית; למשל הרחפת האוטונומי מחליט להתריע לפני המפעיל כי הוא זיהה גורם עוין.

2. שימוש באלגוריתם (K-Nearest Neighbors): KNN כלי זה משמש את המבקר לזיהוי חריגים (Outliers) בסט הנתונים. העיקרון הבסיסי של האלגוריתם הוא שדברים דומים קרובים זה לזה. כאשר האלגוריתם מקבל "דוגמה" חדשה (נקודת נתונים ללא תווית) הוא בודק מי הן הנקודות הקרובות אליה ביותר מתוך הנתונים הקיימים. זיהוי אשכולות נתונים המרוחקים

Russell L. Ackoff, From Data to Wisdom, Ackoff's Best 170 (1999). 25

26 Light Detection and Ranging - גילוי וטווח באמצעות אור.

לאימוץ מדדים רלוונטיים בהיבטי ביצוע וסיכון (KPIs/ KRI²⁷):

א. Model Drift Rate: קצב הירידה בדיוק של מודל ה-AI לאורך זמן.

ב. Poisoning Detection Lag: הזמן החולף מרגע הזרקת נתון זדוני ועד לזיהויו.

ג. Adversarial Robustness Score: עמידות מודל ה-AI מול קליטים מניפולטיביים בבדיקות Red Teaming²⁸.

ד. Mission Accomplishment Under Jamming: שיעור השלמת המשימה בתנאי שיבוש אלקטרוני (אחוז).

ה. Explainability Index: רמת היכולת לנמק החלטות אלגוריתמיות קריטיות.

6. ביקורת מתמשכת (Continuous Auditing): ניטור רציף באמצעות סקריפטים אוטומטיים שמאפשר למבקר לזהות חריגות בזמן אמת, ולא בדיעבד, דבר בעל תרומה ניכרת למניעת כשלים קטלניים במערכות ביטחוניות בלתי מאוישות משולבות AI.

יישום המסגרת: מתודולוגיית ביקורת עדכנית והמלצות לשיפור מתמשך

מסגרת הביקורת המוצעת במאמר מבקשת לחולל שינוי ממוקד אך מהותי באופן שבו ביקורת בתחום הגנת הסייבר ניגשת למערכות ביטחוניות בלתי מאוישות מבוססות AI: מעבר מביקורת "ציות לבקורות" לביקורת "יכולת הוכחה" (Audit-Ready).

המשמעות המעשית היא שכל טענה על אבטחה, על אמינות, על חוסן או על כשירות מבצעית של מערכת ביטחונית בלתי מאוישת מבוססת AI מוכרחה להיתמך בראיות הניתנות לשחזור: מפת תלזיות מלאה²⁹ (של חיישנים-תקשורת-מחשוב-מודל-ממשק החלטה), תיעוד מקור והיסטוריה של נתונים (provenance/תיוג/גרסאות), רישום גרסאות מודל ותהליך שינוי מבוקר, דוחות בדיקה ומדדים (כולל drift ושגיאות קריטיות), יעדי שירות לזמינות נתונים בזמן אמת ותיעוד תרגילי שרידות (resilience drills), וכן ניהול שרשרת

אספקה של רכיבי AI, בדיקות קבלה לספקים ואימות רכיבים. כאשר המסגרת המתודולוגית של הביקורת מיושמת כך היא מאפשרת לביקורת לנסח ממצאים בני הגנה: ממצא המבוסס על ראיה, מסקנה המבוססת על ניתוח סיכון והמלצה מדידה הכוללת בעל תפקיד, לוח זמנים ומדד הצלחה. בכך מתורגמת מורכבות ה-AI לשפת ביקורת ברורה, שאינה תלויה בטכנולוגיה מסוימת, אלא בתקן מזערי של שליטה, שקיפות ושחזור.

מבחינה מוסדית, יישום המסגרת מחייב עדכון ממוקד של התקינה והמתודולוגיה הקיימות; כאמור לא כתחליף לתקנים הקלאסיים אלא כתוספת שכבת AI-Assurance.

בהקשר הישראלי אפשר למנף את תפיסת ההגנה של מערך הסייבר הלאומי כעוגן המתחבר ישירות לקריטריונים ולמדדים של Audit-Ready, וליצור "שפה משותפת" בין הנחיה להגנה ולביקורת.

במשרד מבקר המדינה ההטמעה המעשית צריכה לכלול: הוספת פרק AI-Audit למסמך המתודולוגיה; קביעת "סף מוכנות לביקורת" שבו היעדר ראיות מזעריות הוא ממצא מהותי; ותוכנית דגימות מחייבת הכוללת לפחות מודל אחד, מאגר נתונים אחד ורכיב ספק אחד בכל ביקורת רלוונטית, לצד בדיקת שחזור מוגבלת (Reproducibility) ותרגיל זמינות/שרידות אחד המבוסס ראיות.

כך תוכל הביקורת להעריך לא רק "אם יש בקורות", אלא אם הארגון מסוגל להוכיח שליטה בסיכון AI באופן עקיב לאורך זמן - וזוהי נקודת המפתח לשמירה על כשירות, על בטיחות ועל אמון ציבורי במערכות אוטונומיות בסביבה ביטחונית.

לסיכום אופן יישום המסגרת המאמר מציע להתאים את כלי הביקורת הקיימים באופן הבא:

1. הטמעת מודל "הבטחת משימה" כתקן מחייב לביקורת לכל מערכת ביטחונית בלתי מאוישת.
2. הקמת "סל ראיות דיגיטלי" מחייב לספקים, הכולל יומני למידה ותיעוד ארכיטקטורה.
3. ביצוע מבחני חוסן (Red Teaming) ייעודיים ל-AI המדמים תוקף אדפטיבי.
4. אימוץ כלי הסברתיות (Shapley) וזיהוי חריגים (KNN) כחלק בלתי נפרד מדוחות הביקורת.

27 שני כלי ניהולי משלימים: KPI; Key Performance Indicators, שמטרתו למדוד התקדמות לעבר השגת יעדים; KRI; Key Risk Indicators, שמטרתו לזהות ולנהל סיכונים פוטנציאליים.

28 Red Teaming - צוות אדם - מונח מעולם הביטחון, המודיעין ואבטחת המידע, המתאר תרגול של תקיפה מבוקרת על ארגון כדי לבדוק את מוכנותו. הרעיון העיקרי הוא לאמץ "ראש של יריב" כדי לאתר פרוצדורות, כשלים לוגיים או נקודות תורפה שהארגון עצמו לא הבחין בהם.

29 מפת תלזיות מלאה (Dependency Map או Dependency Matrix) - כלי ויזואלי או לוח המשמש בניהול פרויקטים, בהנדסת מערכות ובפיתוח תוכנה. מטרת המפה היא לנתח ולנהל את כל הקשרים והאינטראקציות בין רכיבים שונים במערכת.

5. הקמת צוותי ביקורת רב-תחומיים, שכן חוסר השקיפות המלא של מודלי ה-AI הוא "בעיה מורכבת" שאינה ניתנת לפתרון בדרכים טכניות בלבד, אלא דורשת שילוב של מומחי סייבר, מדעני נתונים, משפטנים ומבקרים.

סיכום

במערכות ביטחוניות בלתי מאוישות בעידן ה-AI שאלת הביקורת אינה רק אם יש בקרות, אלא אם הגוף המבוקר מסוגל להוכיח, בראיות ניתנות לשחזור, שהמערכת בכללותה - רכיביה והתהליכים המבצעיים המתחוללים בה על בסיס שימוש במודלי AI - שומרת על הבטחת משימה, על חוסן מבצעי ועל אחריותיות לאורך מחזור חייה.

המאמר מציג מסגרת של הבטחת מוכנות לביקורת (Audit-Ready Assurance Framework) בתחום ההגנה מכני איומי סייבר במערכות ביטחוניות בלתי מאוישות משולבות AI, המבוססת על מיפוי התקינה הקיימת והרחבתה לעידן שבו נתונים ומודלי AI הם רכיבי סיכון מרכזיים. המסגרת מגדירה קריטריוני ביקורת ברורים, מדדי ביצוע וסיכון (KPIs/KRIs), סט ראיות מזערי, לצד תוכנית בדיקות ודגימות ישימה לביקורת.

נוסף על כך, המאמר מציע עדכון תפיסה ותקן המחבר בין עקרונות ההגנה של מערך הסייבר הלאומי ובין שיטות הביקורת של משרד מבקר המדינה ומתרגם אותם להמלצות יישומיות לעדכון מתודולוגיית הביקורת. התוצאה היא מעבר מביקורת של "ציות לבקרות" לביקורת של "כולת הוכחה", המאפשרת ממצאים מדידים, השוואת ביצועים וניתוח תוצאות התנהגות של מודלי AI.

נספח: תקנים מרכזיים המקובלים להתוויית המתודולוגיה לביצוע ביקורות בתחום אבטחת המידע והסייבר בארגונים

א. NIST³⁰ CSF³¹ 2.0 (2024): מסגרת תוצרים לניהול סיכון סייבר לכל ארגון; מחזקת מינהל תקין

(Governance), ניהול סיכונים ובקרה על שרשרת אספקה.

ב. NIST SP 800-37 Rev.2 Risk Management Framework (2018): תהליך מחזור חיים לניהול סיכוני אבטחה/פרטיות: אפיון מערכת, בחירת בקרות, יישום, הערכה, אישור וניטור רציף; בסיס מתודולוגי לביקורת ראייתית.

ג. NIST SP 800-53 Rev.5 (2020): קטלוג בקרות אבטחה/פרטיות המאפשר מיפוי בקרות לממצאים; ומספק שפה לביקורת בתחום המינהל התקין - ניהול ותקינה, הרשאות, לוגים, הגנת שרשרת אספקה ועוד.

ד. NIST SP 800-53A Rev.5 (2022): מתודולוגיית הערכת בקרות (בחינה/תצפית/ריאיון) לבניית תוכנית בדיקות ודגימות; תומכת בראיות הניתנות לשחזור.

ה. ת"י (תקן ישראלי) ISO³²/IEC³³ 27001:2022: תקן מערכת ניהול אבטחת מידע (ISMS³⁴) המעוגן בישראל כתקן ישראלי; מחייב מדיניות, בקרות, ביקורת פנימית, טיפול בסיכונים ושיפור מתמיד.

ו. ת"י (תקן ישראלי): ISO/IEC 42001:2023 - מערכת לניהול AI (AIMS³⁵): תקן מערכת ניהול ל-AI; מגדיר דרישות ארגוניות לניהול AI - מינהל תקין, אחריות, תיעוד, בקרת שינוי, ניהול השפעות ומדדים.

ז. The Institute of Internal Auditors (IIA), Global Technology Audit Guide (GTAG) Auditing Cyber Risk in the Age of AI (2020): מדריך ביקורת טכנולוגיה גלובלי של לשכת המבקרים הפנימיים העולמית (IIA), הכולל הנחיות למבקרים פנימיים, למנהלי סיכונים ולחברי דירקטוריון. המדריך הספציפי משנת 2020 מתמקד בשינוי הפרדיגמה של ביקורת בתחום אבטחת הסייבר בעידן ה-AI.

National Institute of Standards and Technology; NIST 30

Cybersecurity Framework; CSF 31

International Organization for Standardization; ISO 32

International Electrotechnical Commission; IEC 33

Information Security Management System; ISMS 34

Artificial Intelligence Management System; AIMS 35

20

המבקר האחרון? בינה מלאכותית ועתיד מקצוע הביקורת

פבל קריינין

מר פבל קריינין הוא מנכ"ל A-Impact, חברה המתמחה בתיווך בין מומחיות אנשית לטכנולוגיית בינה מלאכותית, ומלווה ארגונים במגזר הציבורי והעסקי בתהליכי הטמעת AI.

המאמר בקצרה

- 57% משעות העבודה בכלכלה האמריקאית ניתנות כבר היום לאוטומציה - 44% על ידי סוכנים דיגיטליים ו-13% על ידי רובוטים פיזיים. יכולות סוכני הבינה המלאכותית מוכפלות בכל כמה חודשים.
- השאלה אם בינה מלאכותית עתידה להחליף את המבקר האנושי חדלה להיות ספקולציה עתידית והפכה לשאלה מעשית שמוסדות הביקורת בעולם כבר מתמודדים עימה. מאמר זה מפרק את תהליך הביקורת לרכיביו, בוחן את יכולת הבינה המלאכותית לבצע כל אחד מהם ומציג תרחישי שימוש ממוסדות ביקורת מובילים בעולם - GAO האמריקאי, NAO הבריטי ו-ANAO האוסטרלי.
- ייחודו של המאמר אינו במיפוי הטכנולוגיה, כי אם בהיפוך שאלת המוצא. הדיון הציבורי עוסק בשאלה כיצד תוכל הבינה המלאכותית להשתלב לצד המבקר האנושי ולהעצים את עבודתו. המאמר טוען שזו שאלה זהירה מדי, כמעט מתחמקת.
- השאלה הנוקבת היא אחרת: האם יש בכלל הצדקה להותיר את הביקורת בידיים אנושיות? כיום, חלקים מן המגזר הציבורי אינם מבוקרים כלל. הדגימה מצומצמת, הבדיקה נעשית רק בדיעבד, והדוחות מתפרסמים לעיתים שנים אחרי האירוע. יוצא אפוא שגם המבקר המיומן ביותר סורק רק חלק זעיר ממה שראוי להיבדק. מכונה יכולה לשנות כל אחד מהפרמטרים הללו - היקף, תדירות וזמן תגובה: מדגימה אל כיסוי מלא, מרטרוספקטיבה אל ביקורת רציפה ומדוחות שמגיעים לאחר שנים אל תוצר שמגיע בתוך ימים.
- אם תכלית הביקורת היא לשרת את הציבור, הרי שההיצמדות לגורם האנושי עלולה להתגלות לא כנאמנות לערך הביקורת אלא כפגיעה בו.
- מוסדות הביקורת שיקדימו להיערך לא רק ישמרו על הרלוונטיות שלהם - הם יגדירו מחדש את הסטנדרט שלפיו תימדד ביקורת המדינה בעשורים הבאים.

מבוא | מיפוי פעולות הביקורת ובחינתן ביחס ליכולת הבינה המלאכותית | יישומי הבינה המלאכותית כיום | תרחישי שימושי AI בעבודת הביקורת בפועל | ביקורת על גופים מבוקרים המשתמשים בבינה מלאכותית | מצ'ט לסוכן: רמת האוטונומיה של הכלי | התהליך מול התוצר: האינטרס הציבורי בקיום ביקורת אנושית | האם הבינה המלאכותית תייתר את עובדי הביקורת? | סיכונים ואתגרים | ממתודולוגיה ליישום: מפת דרכים | מתווה אופרטיבי לשילוב בינה מלאכותית בעבודת הביקורת | סיכום

מבוא - העין הפקוחה

נוספות ניתנות לאוטומציה על ידי רובוטים פיזיים. אלו אינן שתי סטטיסטיקות נפרדות, אלא שני חלקים של אחד, המשקפים יחד את הפוטנציאל הכולל של האוטומציה.

התמונה אינה שונה כשבוחנים את עולם הביקורת. זוהי גם תמונת המצב שעולה מדוח Anthropic⁵ ממרץ 2026 על השפעות AI על שוק העבודה⁶. על פי דוח זה, תפקידי משרד ומינהלה, מקצועות עסקיים-פיננסיים ומקצועות משפטיים - שאליהם משתייכים גם תחומי הביקורת והחשבונאות - משתייכים לקטגוריות שלהן חשיפה תיאורטית של 80% - 94%. כך עולה גם מדוח IIA Global Internal Audit Standards 2024⁷, הטוען כי תפקידים הכרוכים בניתוח מידע, הערכת ציות, בדיקת נתונים כספיים ועריכת דוחות נמנים עם המקצועות בעלי החשיפה הגבוהה ביותר לאוטומציה. אלה בדיוק היכולות שדורשת ליבת עבודת המבקר.

בד בבד, המגזר הציבורי, בעיקר בארצות הברית, בבריטניה ובאוסטרליה, כבר מאמץ בינה מלאכותית בקצב מואץ. סקר של ה-National Audit Office הבריטי (NAO) משנת 2024 מצא כי 37% מגופי הממשלה כבר פרסו מערכות AI, ו-70% נוספים מתכננים לעשות כך או מפתחים יישומים⁸. בארצות הברית קפץ מספר תרחישי השימוש ב-AI בממשל הפדרלי מ-2,133 ב-2024 ל-3,611 ב-2025 - גידול של כ-70% בשנה אחת בלבד, שבמהלכה הצטרפו 15 סוכנויות נוספות ומספרן הכולל עלה ל-56. מתוך 3,611 התרחישים, 1,818 כבר פרוסים או בפיילוט ו-445 מסווגים כ'השפעה

מהות הביקורת - הן ביקורת המדינה והן הביקורת הפנימית - היא "העין הפקוחה": מנגנון שנועד להבטיח לציבור שמי שמנהל את ענייני עושה זאת כראוי, ביעילות, בחיסכון ובאפקטיביות. בהיסטוריה האנושית המוכרת, העין הזו הייתה תמיד אנושית. מבקרים אנושיים נכנסו לגופים מבוקרים, בחנו מסמכים, ראיינו עובדים, ניתחו נתונים, ניסחו ממצאים והגישו דוחות. מערכת שלמה של ידע, ניסיון ותפיסת עולם מקצועית נבנתה סביב התהליך הזה. השאלה שמאמר זה מעלה אינה אם העין הזו חשובה - היא חשובה מתמיד - אלא אם היא חייבת להיות רק אנושית. המסגרת הנורמטיבית של הביקורת הציבורית מעוגנת בתקן ISSAI 100 של INTOSAI¹ בחוק-יסוד: מבקר המדינה (ס' 2) ובחוק הביקורת הפנימית, התשנ"ב-1992.

מכון המחקר העולמי של מקינזי - MGI (McKinsey Global Institute)² פרסם בנובמבר 2025 דוח שטלטל את עולם התעסוקה וקבע כי 57% משעות העבודה במקצועות הקוגניטיביים והמשרדיים בכלכלה האמריקאית ניתנות לאוטומציה באמצעות הטכנולוגיה הקיימת היום³. אין מדובר בתחזית עתידית או בהערכה תיאורטית. מדובר ביכולת טכנית שכבר קיימת, כאן ועכשיו. אך כדי להבין נכון את הנתון הזה, יש להביט בפילוח שמציג הדוח⁴: 44% משעות העבודה ניתנות לאוטומציה על ידי סוכני בינה מלאכותית (להלן גם - AI) - תוכנות אוטונומיות שמבצעות משימות קוגניטיביות ודיגיטליות. 13% שעות

1 International Organization of Supreme Audit Institution (INTOSAI), ISSAI 100 - Fundamental Principles of Public-sector Auditing (2019), https://www.intosai.org/fileadmin/downloads/documents/open_access/ISSAI_100_to_400/issai_100/ISSAI_100_EN.pdf תקן בין-לאומי המגדיר את עקרונות היסוד של הביקורת במגזר הציבורי כביקורת עצמאית, אובייקטיבית, ממלכתית ומלווה באחריותיות.

2 דו"ע המחקר של חברת הייעוץ McKinsey & Company שפרסומיה משרתים מקבלי החלטות במגזר הציבורי ובמגזר הפרטי.

3 McKinsey Global Institute, *Agents, Robots, and Us: Skill Partnerships in the Age of AI* (2025), <https://www.mckinsey.com/mgi/our-research/agents-robots-and-us-skill-partnerships-in-the-age-of-ai>

4 שם, בעמ' 14. הנתון מתייחס ליכולת הטכנית בלבד ולא לתחזית ההחלפה בפועל.

5 Anthropic היא חברת בינה מלאכותית מובילה, מפתחת משפחת מודלי השפה Claude.

6 Anthropic, *Labor Market Impacts of AI: A New Measure and Early Evidence* (2026), <https://cdn.sanity.io/files/4zrzovbb/chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://cdn.sanity.io/files/4zrzovbb/website/2b5bbaf2c1eb81dbf6e6fb813c1a24e35a64d376.pdf>

7 The Institute of Internal Auditors (IIA), *Global Internal Audit Standards* (2024), theiia.org/en/standards/2024-standards/global-internal-audit-standards אפקטיבי מ-9 בינואר 2025.

8 National Audit Office (UK), *Use of Artificial Intelligence in Government*, HC 612 (2024), nao.org.uk/reports/use-of-ai-in-government. הסקר נערך בקרב 87 גופי ממשלה בבריטניה. ראוגם: House of Commons Committee of Public Accounts, *Use of AI in Government*, HC 356 (2025), <https://publications.parliament.uk/pa/cm5901/cmselect/cmpubacc/356/report.html>

לשינוי אלא אף להוביל אותו. מוסד ביקורת המדינה הוא בעל סמכות רחבה ויכולת השפעה על הממשל כולו ויכול להיות חלוץ בתחום זה. משכך, הרלוונטיות של הדיון הזה לישראל היא מיידית ומעשית.

מאמר זה מבקש לבחון את הטענה שעבודת הביקורת ניתנת לאוטומציה - לא כספקולציה עתידנית, אלא כמציאות טכנולוגית שמתהווה מול עינינו ושראוי כי אנשי מקצוע בתחום הביקורת יכירו אותה לעומקה. המאמר יפרק את תהליך הביקורת לרכיביו, יבחן את מצב הטכנולוגיה, יציג תרחישי שימוש אמיתיים ויטען כי דווקא בתחום הביקורת שימור האדם במוקד התהליך עלול להיות מנוגד לערך שהביקורת אמורה לשרת ולא לשמור עליו.

מה באמת עושה מבקר? מיכיו פעולות הביקורת ובחינתן ביחס ליכולות הבינה המלאכותית

תהליך הביקורת - הן ביקורת המדינה והן ביקורת פנימית - ניתן לפירוק לתשעה רכיבי יסוד המשקפים את מבנה תהליך הביקורת כפי שהוא מעוגן בתקני INTOSAI ISSAI (Guidelines on Central Concepts for Performance Auditing) ובתקני IIA Global Internal Audit Standards (2024), הכוללים שלבי תכנון, ביצוע, דיווח ומעקב. כל רכיב מייצג סוג שונה של פעילות אינטלקטואלית, וניתן לבחון את יכולתה של הבינה המלאכותית לבצע כל אחד מהם בנפרד.

תכנון

שלב התכנון מתחיל בהערכה עמוקה של הנושא המבוקר: מהו תחום הפעילות, מהם הסיכונים הטבועים בו, איפה נדרשת העמקה מיוחדת והיכן יש להשקיע את משאבי הביקורת. זהו שלב שדורש הבנה של סביבה רגולטורית, זיהוי רבדים טעונים (חוקיים, פוליטיים וחברתיים) ותיעודף רציונלי של המשאבים המוגבלים של צוות הביקורת.

גבוהה⁹. המשמעות היא שהגופים שמסודות הביקורת אמורים לבקר הופכים בעצמם לגופים מונעי AI. מוסד ביקורת שאינו שולט בטכנולוגיה הזו לא יוכל לבקר את מי שמשתמש בה, שכן לא יוכל להעריך את אמינות התוצרים של המערכות המבוקרות, לאתר הטיות אלגוריתמיות או לזהות סיכונים ממשי AI ייחודיים.

ישראל בעידן הבינה המלאכותית

גם בישראל קיימת מודעות גוברת לחשיבות הנושא. במשרד מבקר המדינה נוציב תלונות הציבור הוקמו אנשים ייעודיים לביקורת הבודקת טכנולוגיות דיגיטליות וסייבר. מבקר המדינה המכהן ר"ח מתניהו אנגלמן, מגלה עדיפות לביקורת טכנולוגית בתיאום בין-לאומי¹⁰. בדוח מבקר המדינה משנת 2024 בנושא המוכנות של ישראל לעידן הבינה המלאכותית נמצאו ליקויים ניכרים בנושא זה, בין היתר היעדר גורם מרכזי המוביל את אסטרטגיית ה-AI בממשל, חוסר תיאום בין יחידות הממשלה וחוסר בהירות במדיניות.

עדות למיקומה המרכזי של ישראל בשילוב AI בעולם הביקורת ניתן לראות בין היתר בכנס בין-לאומי שהתקיים באבו דאבי בפברואר 2026 בהובלת מבקר המדינה בתפקידו כנשיא EUROSAI. בכנס הוצג דוח ביקורת בין-לאומי בנושא AI שהכין משרד מבקר המדינה בשיתוף 11 משרדי ביקורת עליונים אירופיים נוספים. על המאמר היו חתומים, בין היתר, מזכ"ל ה-OECD מתיאס קורמן ומבקר המדינה של איחוד האמירויות חומיד עובד אבו-שיבס¹¹.

ישראל היא מדינה המתהדרת בחדשנות טכנולוגית. כך, לפי מדד Tortoise AI Global Index 2024, ישראל מדורגת במקום ה-9 מבין 83 מדינות, היא בעלת אקו-סיסטם עשיר בחברות הזנק (סטארט-אפ) ונהנית מנוכחות טכנולוגית גבוהה גם בעולמות ה-AI (לפי דוח Startup Nation Central לשנת 2024, בישראל מעל 2,500 חברות AI פעילות; חלקן של חברות אלה בגיוסי הון עמד על כ-30% מסך גיוסי הטכנולוגיה). כלומר, ביכולתה של ישראל לא רק להגיב

9 U.S. Office of Management and Budget (OMB), 2025 Federal Agency Artificial Intelligence Use Case Inventory (2025), github.com/ombegov/2025-Federal-Agency-AI-Use-Case-Inventory

10 משרד מבקר המדינה דוח ביקורת מיוחד - ההערכות הלאומית בתחום הבינה המלאכותית (2024) EUROSAI IT Working, library.mevaker.gov.il/sites/DigitalLibrary/Documents/2024/2024.11-Cyber/2024.11-Cyber-107-AI.pdf; Group, ITWG Newsletter (2024), <https://euosai-it.org/news/newsletter>. (משפט כלכלי) עקרונות, מדיניות, רגולציה ואתיקה בתחום הבינה המלאכותית 2023 (2023), https://www.gov.il/he/pages/ai_23

11 משרד מבקר המדינה נוציב תלונות הציבור, "מזכ"ל ה-OECD למבקר המדינה: 'הממשלות חייבות לאמץ את יתרונות הבינה המלאכותית ולמזער את הסיכונים'" (3 בפברואר 2026), www.mevaker.gov.il/newsroom/2026-02-03-eurosai

12 International Organization of Supreme Audit Institution (INTOSAI), ISSAI 3100: Standard for Performance Auditing (n.d.), <https://www.issai.org/pronouncements/issai-3100-performance-audit-guidelines-key-principles/>. התקן מגדיר עקרונות לביקורת ביצועים שניתנים ליישום גם בהקשר של מערכות AI.

מסוגל ללמוד במלואו ולעומקו בתוך זמן קצוב. זהו שלב בעבודת הביקורת שכבר התחוללה בו תמורה בשנתיים האחרונות: ככל שגופים מבוקרים עוברים לדיגיטציה, המידע הופך נגיש יותר לכלים אוטומטיים. סוכן AI המבוסס על מסגרת כמו Microsoft AutoGen¹⁸ או LangChain Agents ומחובר למערכות המידע של הגוף המבוקר יכול לשאוב חומרים בהיקף ובמהירות שאין להם תקדים - אלפי מסמכים, מיליוני רשומות ושנים של נתונים בתוך שעות ספורות בלבד.

קביעת סרגל

הסרגל - הנורמה שעליה יתבססו הממצאים - הוא התשתית לקביעת ממצאי הביקורת, והוא עשוי לכלול מצע קשיח של חיקוקים, פסקי דין והחלטות שיפוטיות, נהלים, הוראות רגולטוריות, הנחיות מדיניות ותקנים מקצועיים בין-לאומיים. התשתית האמורה עשויה לכלול גם נורמות רכות, בין היתר מספרות המחקר ומדוחות מקצועיים. קביעת הסרגלים ויישומם דורשים בקיאות במקורות ובתימוכין ולימוד תכנים רלוונטיים לנושא הנבדק באופן מעמיק.

במסגרת ביצוע שלב זה בעבודת הביקורת עשוי להיות לבניה המלאכותית יתרון מוכח על הבניה האנושית. מודלי שפה גדולים כגון Claude¹⁹ Opus 4.7, Gemini 3.1²⁰ Pro ו-GPT-5.5²¹, מסוגלים לסרוק, לנתח ולהצליב מאות מסמכים נורמטיביים - חקיקה ראשית ומשנית, פסיקה, נהלים פנימיים ותקנים בין-לאומיים כדוגמת תקני ISSAI²² - ולהפיק מהם סרגל ביקורת מדויק ומעודכן בתוך דקות. יתרה מכך, הכלי יכול להתעדכן בזמן אמת בשינויי חקיקה ורגולציה, יכולת שמבקר אנושי לא תמיד נהנה ממנה.

כלי AI מתקדמים כבר מסוגלים לנתח מאגרי מידע נרחבים, לזהות דפוסי סיכון ולהציע תיעודי מבוסס נתונים. כך למשל, Thomson Reuters CoCounsel Audit מקצר תהליכי סקירת workpapers בעד 70%,¹³ MindBridge Ai Auditor מאפשר בדיקה של 100% מהעסקאות בעת ניתוח סיכונים,¹⁴ AuditBoard/Optro הוא סוכן בדיקה לבקורות פנימיות אשר משמש יותר מ-50% מחברות Fortune 500 ו-Microsoft 365 Copilot¹⁵ מציע תסריטים מוכנים לשימוש בתהליכי הביקורת.^{17,16} לעיתים קרובות כלים אלה מציעים כיסוי רחב יותר מזה שצוות אנושי יכול להשיג - כפי שהוכיח ה-GAO עצמו במערכת פנימית שפיתח לצורך סינתזה של דוחות ביקורת קודמים ולסריקת מסמכים של הקונגרס. מערכת AI כגון MindBridge Ai Auditor, שעברה תיקוף על ידי UCLC ועומדת בדרישות תקני ISA-99, SAS 240, CAS 240, יכולה למשל לסרוק את כל דוחות הביקורת מהשנים האחרונות, לזהות נושאים חוזרים שלא טופלו ולהציע תיעודי אוטומטי על בסיס תדירות, חומרה וציר זמן. כך למשל פועל סוכן הבדיקה של AuditBoard/Optro.

איסוף חומרים

הגישה לחומרים המצויים ברשות הגופים המבוקרים היא אתגר מקצועי ראשון במעלה עבור עובדי הביקורת. מדובר בחומר כתוב (מסמכים אקטואליים או ארכיונים) המפוזר במקומות אחסון מגוונים ובמערכות מידע שלא תמיד מכילות רכיבי תיעוד ויצוא נתונים נוחים ונגישים לשימוש בעבודת הביקורת. כדי להשיג את החומרים נשואי הביקורת, על המבקר להשיג הרשאות גישה, לנווט במערכות שונות ולהתמודד עם היקף של מידע שלעיתים צוות אנושי אינו

Thomson Reuters, CoCounsel Audit - Legal Research and Audit Workpapers, <https://tax.thomsonreuters.com/en/products/cocounsel-audit> 13

MindBridge AI, Platform Overview - Finance-Native AI Trained on 260+ Billion Transactions, mindbridge.ai 14

Optro (formerly AuditBoard), AI Platform for Internal Controls (2026), optro.ai 15

Microsoft, Microsoft 365 Copilot - Audit and Risk Use Cases, adoption.microsoft.com/en-us/scenario-library/ 16

U.S. Government Accountability Office (GAO), Artificial Intelligence: GAO's Work to Leverage Technology and Ensure Responsible Use 17
GAO-24-107237 (2024), gao.gov/products/gao-24-107237. ה-GAO מתאר במסמך את פריסת מודל השפה הגדול ככלי לסינתזה של דוחות עבר וסריקת מסמכי קונגרס.

Microsoft Research, AutoGen: Multi-Agent Conversation Framework, microsoft.github.io/autogen 18

Anthropic, Claude Opus 4.7 - 1M Token Context Window (April 2026), platform.claude.com/docs/build-with-claude/context-windows. 19

Google, Gemini 3.1 Pro - Long Context Window of 1M Tokens (April 2026), ai.google.dev/gemini-api/docs/long-context. 20

OpenAI, Introducing GPT-5.5 (April 23, 2026), openai.com/index/introducing-gpt-5-5. 21
שוחרר ב-23 באפריל 2026 עם חלון הקשר של מיליון טוקנים.

ISSAI (International Standards of Supreme Audit Institutions) הם תקנים בין-לאומיים של INTOSAI (ארגון הגג של מוסדות הביקורת העליונים בעולם) המגדירים את עקרונות הביקורת במגזר הציבורי. מוסדות הביקורת העליונים החברים ב-INTOSAI כפופים לתקנים אלה.

ניתוח - הפער בין המצוי לרצוי

לצורך זיהוי הפער נדרש תחילה שלב מקדים של קביעת המצב המצוי והעובדות - איסוף ראיות, אימות מסמכים וראיונות - כמעגל ב-ISSAI 3100 Performance Audit Guidelines ובכרך התכנון של IIA Global Internal Audit Standards 2024.

השוואה שיטתית בין המצוי (כיצד פועל הגוף בפועל) לבין הרצוי (הנורמה) מצויה בליבת עבודת הביקורת. הסוכן - המבקר - צריך לאסוף נתונים על מצב העניינים בפועל (דוחות פיננסיים, מדיניות בפועל, החלטות מינהליות והיבטי ביצוע) ולהשוות אותם בשיטתיות לרף ההתנהלות שהציבה הנורמה. זוהי משימה שדורשת זיהוי פערים, הערכה של חשיבותם והשפעתם וניתוח של ההשלכות. תהליך זה מעוגן בתקני IIA (5 תחומים, 15 עקרונות מנחים, 52 תקנים ומעל 200 דרישות) ובתקני ISSAI 3100.²³

כלי בינה מלאכותית כגון Thomson Reuters CoCounsel, MindBridge Ai Auditor, AuditBoard/Optro, Microsoft 365 Copilot ואחרים מסוגלים לצלוח את השלב הזה באופן מיטבי, לרבות השוואת מערכי נתונים גדולים (Big Data), זיהוי חריגות וסטיות, סיווג פערים לפי חומרה ודחיפות ותייעוד שיטתי של הממצאים. סוכן AI מסוגל עקרונית לבדוק כל עסקה, כל חוזה וכל החלטה - לא מדגם, אלא 100% מהאוכלוסייה. יכולת כזו יש לדוגמה למערכת MindBridge, אשר מאומנת על יותר מ-260 מיליארד עסקאות ביותר מ-3,000 מערכות ERP, עם יותר מ-8,000 חוקי GAAP מוטמעים.²⁴ בעת השימוש במערכת הבקרה האנושית נדרשת בעיקר לשם אימות חריגות שהכלי מסמן לבדיקה.

גיבוש ממצאים, מסקנות והמלצות

המדריך לביקורת המדינה²⁵ מגדיר ממצא ביקורת כ"מידע עובדתי רלוונטי הנובע מהשוואה בין הנורמה למצב בפועל, כפי שעולה מראיות הביקורת". ממצא המבטא פער שלילי ביחס לנורמה מחייבת הוא ליקוי. המדריך מבחין בין שני שלבים: השלב הראשון הוא גיבוש הממצא - שלב שבו אוספים את הראיות, בוחנים את ההשוואה לנורמה, את השפעות הליקוי לטווח הארוך ואת השלכות הרוחב שלו

ובודקים אם יש אינדיקציה לבעיה שורשית ומהי חומרת התמשכותו של הליקוי. השלב השני הוא ניסוח הממצא בדוח - תרגום הקביעה העובדתית לטקסט ברור, מאוזן, מנומק ומבוסס ראיות, המשקף את עמדת מוסד הביקורת בלשון מדויקת ובהירה לקורא הסביר. כל אחד מהשלבים דורש כישורים מקצועיים נפרדים: הראשון - שיקול דעת ביקורתי וניתוח עובדתי, השני - מיומנויות ניסוח, עריכה ובהירות. בשלב הניסוח ניתן להיעזר במודלי שפה גדולים כעורכי משנה: יצירת טיוטה ראשונית מתוך הממצאים שגובשו, התאמת רמת הכירוט לקהל היעד, בדיקת עקביות מול דוחות קודמים ואיתור עמימות לשונית. על המבקר האנושי יוטל לשמור על שיקול הדעת המקצועי ועל האחריות לתוכן.²⁶

מודלי שפה גדולים כגון Claude Opus 4.7, GPT-5.5 ו-Gemini 3.1 Pro הם כותבים מיומנים. הם מסוגלים לנסח ממצאים בהתאם לפורמט קבוע, לנמק את האמור על בסיס הנורמה שהופרה, להצליב נתונים עם ממצאים דומים מביקורות קודמות ולהציע המלצות מותאמות. כך, המבקר האנושי יעבור מתפקיד כותב לתפקיד עורך - שינוי שיחסוך זמן רב וישפר את עקביות הדוחות. כלומר, המבקר האנושי יבחן, יאמת וישפר טיוטה שחיברה מערכת AI, וכך יתפנה להתמקד בשיקול הדעת המקצועי ובקבלת ההחלטות, במקום להשקיע במלאכת הניסוח עצמה.

שלב הפנייה - קבלת תשובות הגופים המבוקרים והתייחסות אליהן

לאחר השלמת טיוטת הדוח, המבקר שולח אותה לעיון הגוף המבוקר ולמתן הערות על תוכן הטייטה וסגנונה. הגוף המבוקר מגיב לאמור בדוח, מתייחס לתכנים ומעיר עליהם, ובמקרים רלוונטיים לשיטתו מנסה להוכיח שהתמונה העובדתית שהציגה הביקורת אינה הולמת את המציאות בפועל וממצא הביקורת שגוי, אינו מדויק או חסר ביסוס. במקרים מסוימים עשוי הגוף המבוקר לטעון כי הגם שממצאי הביקורת נכונים לכאורה, הרי שהערות הביקורת והמלצותיה אינן רלוונטיות או שיישומן בפועל אינו מעשי. בהתייחסו לתגובת הגוף המבוקר, על המבקר להעריך את היקף טענותיו ואת מידת נכונותן, מהימנותן, תקפותן והשפעתן על נוסח ממצאי הדוח.

The Institute of Internal Auditors, Global Internal Audit Standards 2024, <https://www.theiia.org/en/standards/2024-standards/global-internal-audit-standards>, אפקטיבי מ-9 בינואר 2025.

MindBridge AI, *Platform Overview - Finance-Native* 24, לעיל ה"ש 15.

25 משרד מבקר המדינה ונציב תלונות הציבור *מדריך לביקורת המדינה* (ללא ציון תאריך), <https://www.mevaker.gov.il/state-audit/roles-work/state-audit-guide>

26 שם. המדריך מגדיר את מתודולוגיית הביקורת הממלכתית ואת שלבי עבודתו של המבקר.

מותאמים לקהלי יעד שונים (תקציר מנהלים בעמוד אחד, תקציר תקשורתי ותקציר מקצועי מעמיק), ייצור פורמטים נגשים (יצירת הסכת [פודקאסט] מהדוח באמצעות מערכות כגון ElevenLabs, הפקת סרטוני וידאו קצרים, הכנת אינפוגרפיקות אינטראקטיביות ועוד), תרגום אוטומטי איכותי לשפות נוספות (ערבית, אנגלית, רוסית ואמהרית) לטובת אוכלוסיות בישראל ובחו"ל וכן ניתוח של הכיסוי התקשורתי ותגובת הציבור ברשתות החברתיות לאחר הפרסום בעזרת כלים כגון Brandwatch ו-Meltwater. הפרסום הוא חלק אינטגרלי מתפקיד הביקורת - דוח שלא נקרא ולא הופנם חוטא לתכליתו הציבורית.²⁸

מעקב ותיקון ליקויים

המעקב אחר יישום המלצות הביקורת הוא תהליך מתמשך ולעיתים סיזיפי. בדרך כלל, מוסדות הביקורת מבצעים בדיקות מעקב ספורדיות על יישום המלצות דוחות הביקורת ותיקון הליקויים שתועדו בהם לאחר פרקי זמן משתנים (החל בשנתיים לאחר פרסום הדוח). אולם עד ביצוע ביקורת המעקב עשויה המציאות בשטח להתעדכן באופן שאינו מתיישב עם ממצאי הדוח המתיישן, ולכן ההמלצות בדוח עלולות להפוך ללא רלוונטיות או שיישומן עלול לפגוע במועילות תפקודו של הגוף המבוקר.

לעומת המעקב האנכרוניסטי, סוכן AI שמחובר למערכות הארגון יכול לבצע מעקב רציף - לא בדיקה תקופתית אחת לשנה, אלא ניטור מתמשך המתריע בזמן אמת על סטיות מתוכנית התיקון. יכולת זו ממומשת כיום בכלים כמו AuditBoard/Optro ו-MindBridge Ai Auditor, המאפשרים ניטור רציף של תהליכים פיננסיים, זיהוי חריגות בזמן אמת והתרעה אוטומטית על סטיות מבקורות מאושרות.²⁹

כפי שעולה מהאמור עד כה וכפי שיורחב להלן - בשנת 2026 סוכני בינה מלאכותית שכבר פותחו בשדה הטכנולוגי הם בעלי יכולת תפעול של כל שלבי עבודת הביקורת, לכאורה ללא מגע יד אדם. לעבודה האנושית תפקיד בניטור כשלים פוטנציאליים ובבקרה על התוצרים, אך גם תפקידים אלה עשויים אולי להתייתר בשנים הקרובות עם התפתחות כלי הבינה המלאכותית.

זהו שלב שמערב אינטראקציה בין אנשים ובין ארגונים, ויש בו כדי לאתגר את הבינה המלאכותית. אך גם בשלב זה ניתן לדמיין מערכת AI שמנהלת את התהליך - גם אם נכון להיום מדובר בתרחיש עתידי המובא כדי להדגים את גבולות הדיון הרגולטורי ולא כיכולת הקיימת באופן מיידי. מערכת כזו תשלח טיטות, תקבל תגובות, תנתח את מידת הרלוונטיות שלהן, תזהה אילו הערות דורשות שינוי מהותי ותנסח את הגרסה הסופית.

ייעוץ מקצועי משלים

לפני סגירת הדוח, המבקר נעזר בייעוץ חיצוני בתחומים הרלוונטיים לממצאים: ייעוץ משפטי לבחינת המסגרת הנורמטיבית והחשיפות, ייעוץ כלכלי וחשבונאי לאומדן השלכות תקציביות, ייעוץ סטטיסטי לאימות מתודולוגיית הדגימה וייעוץ מומחים לנושאים ספציפיים שהדוח עוסק בהם (לדוגמה סייבר, רפואה, הנדסה או סביבה). בשלב זה ל-AI תרומה פוטנציאלית ניכרת: מודלי שפה גדולים מסוגלים לאתר פסיקה וחקיקה רלוונטית בתוך דקות (בכלים כגון Thomson Reuters Westlaw Precision AI או Lexis+), לסכם חוות דעת מומחים מקבילות מתוך מאגרים בין-לאומיים, להציג זוויות מקצועיות נוספות שמבקר אנושי עלול להחמיץ ולהכין מסמכי רקע לדיוני ועדת הביקורת. השלב מסתיים בהטמעת תובנות הייעוץ בדוח עצמו ובחיזוק הביסוס המקצועי של הממצאים.²⁷

השלב האחרון הוא עריכת לשון, עיצוב והכנה לפרסום, תחומים שבהם הבינה המלאכותית כבר פועלת ביעילות רבה ברחבי העולם, מסיכום ועריכה ועד ליצירת מסמכים מובנים בפורמטים מגוונים, כולל גרפיקה, טבלאות ותרשימים.

הפרסום והשיווק של הדוח

לאחר השלמת הדוח הסופי, מוסד הביקורת אחראי לפרסומו ולתרגומו למסרים שיגיעו לקהלים השונים: תקשורת, מקבלי ההחלטות בכנסת ובממשלה, הציבור הרחב, הקהילה המקצועית והגוף המבוקר. תרומת ה-AI בשלב זה יכולה לכלול יצירה אוטומטית של תקצירים

Thomson Reuters, *Westlaw Precision with CoCounsel - AI-Assisted Legal Research*, thomsonreuters.com; LexisNexis, *Lexis+ AI 27* Platform, lexisnexis.com. כלי הייעוץ של AI נסקרים גם ב-*International Bar Association, AI for Legal Professionals: A Practitioner's Guide* (2025), www.ibanet.org/document?id=IBA-Legal-Agenda-2025.

ElevenLabs, *AI Voice Generation Platform*, elevenlabs.io; Brandwatch, *Consumer Intelligence Platform*, brandwatch.com; Meltwater, 28 INTOSAI-P 12, *The Value and Benefits of the Audit*, meltwater.com. על חשיבות הפרסום ככלי ביקורת ראו *Media Intelligence and Social Listening*, meltwater.com. *of Supreme Audit Institutions* (2019), <https://www.issai.org/pronouncements/intosai-p-12-the-value-and-benefits-of-supreme-audit-institutions-making-a-difference-to-the-lives-of-citizens/>.

MindBridge AI, *Continuous Monitoring and Anomaly Detection* (2025), mindbridge.ai/blog/modernizing-internal-controls-over- financial-reporting-with-ai; AuditBoard, *Optro AI Agent Platform*, auditboard.com

יישומי הבינה המלאכותית כיום

כדי להבין את עומק השינוי שעומד בפתח, יש לבחון שלושה מחקרים מרכזיים שפורסמו לאחרונה. כל אחד מהם מאיר היבט אחר של המציאות הטכנולוגית, ויחד הם מציירים תמונה שלא ניתן להתעלם ממנה.

הפוטנציאל הטכני

כאמור, דוח של מכון המחקר העולמי של מקינזי מציג במאמר בשם "Agents, Robots, and Us: Skill Partnerships in the Age of AI" ניתוח שיטתי של פוטנציאל האוטומציה בשוק העבודה האמריקאי. הממצא המרכזי המובא בו הוא כי 57% משעות העבודה ניתנות לאוטומציה כבר באמצעות הטכנולוגיה הקיימת. סוכני AI (תוכנות שמבצעות משימות קוגניטיביות באופן אוטונומי) יוכלו להחליף 44% מהשעות, ורובוטים (מערכות פיזיות) יוכלו להחליף 13% מהן³⁰. כ-43% משעות העבודה דורשות ביצוע משימות פיזיות, אינטראקציה חברתית-רגשית ושיקול דעת מקצועי מורכב, ולפי הדוח הן ימשיכו לדרוש מעורבות אנושית מרכזית בעתיד הנראה לעין. הדוח אומנם מדגיש שמדובר ביכולת טכנית של הבינה המלאכותית ולא בהחלפת עובדים בפועל - אבל הפער בין היכולת לביצוע הולך ומצטמצם עם כל רבעון שחולף.

דוח נוסף של מכון המחקר העולמי של מקינזי, "A New Future of Work: The Race to Deploy AI and Raise Skills in Europe and Beyond", מוסיף נתונים אירופיים ונותן אומדן זמנים לשינוי באירופה: עד 2030, כ-30% משעות העבודה באירופה יעברו אוטומציה, עם צורך ב-12 מיליון שינויי תעסוקה - הכפלה של קצב השינויים לעומת העשור שלפני מגפת הקורונה³¹.

הפער בין הפוטנציאל למציאות

על אף האמור לעיל, תמונת המצב בשטח שונה. בסקר

ה-NAO הבריטי על השימוש ב-AI בממשלה דיווחו 37% מתוך 87 הגופים הממשלתיים שנסקרו כי הם משתמשים ב-AI³², ו-70% מהנתרים היו בשלבי הפיילוט או התכנון³³. בארצות הברית קפץ מספר יישומי ה-AI המשמשים בסוכנויות פדרליות מ-571 ב-2023 ל-1,110 ב-2024³⁴. אם 57% משעות העבודה ניתנות לאוטומציה, מדוע זה לא קורה כבר עכשיו? מחקר פורץ דרך של Anthropic ממרץ 2026, "Labor Market Impacts of AI: A New Measure and Early Evidence", מספק לשאלה זו תשובה מדויקת ומפתיעה. המחקר מציג לראשונה מדד חדש: "חשיפה נצפית" (observed exposure). מדד זה אינו בודק מה הבינה המלאכותית מסוגלת לבצע באופן תיאורטי, אלא מהן יכולותיה האקטואליות, הנקבעות על בסיס ניתוח מיליוני אינטראקציות אמיתיות.

הנתונים חושפים פער דרמטי בין הפוטנציאל למציאות: בתחום המחשוב והמתמטיקה, 94% מהמשימות ישימות תיאורטית בעזרת AI, אך רק 33% מבוצעות כך בפועל. בתחומי המשרד והמינהלה מדובר ב-90% מול כ-25%. בתחומים עסקיים ופיננסיים - 85% מול כ-20% ובתחום המשפטי - 80% מול כ-15%³⁵. הפער הזה אינו חולשה של הטכנולוגיה - הוא חלון זמן. חלון זה הולך ונסגר ככל שהכלים משתפרים וככל שהמחירים יורדים כתוצאה מעלייה ביעילות העיבוד של מודלי השפה לצד ירידה בעלות הקריאה ל-API. כך למשל, מחיר עלות שאילתה אחת מהמודל, הנמדד לפי מספר הטוקנים שעוברים בקלט ובפלט של מודלי שפה, ירד ביותר מ-90% במהלך 2023 - 2024.

מחקרים אחרונים מצביעים על כמה גורמים עיקריים לאי-הטמעת כלי AI בפרקטיקה הארגונית המצויה אף שהטכנולוגיה מאפשרת זאת:

א. חשש מחוסר דיוק - ארגונים רבים מדווחים על חשש מחוסר דיוק אשר מעכב הטמעת AI. אם AI עלול לטעות, גם

30 שם; McKinsey Global Institute, *Agents, Robots, and Us*, לעיל ה"ש 3.

31 McKinsey Global Institute, *A New Future of Work: The Race to Deploy AI and Raise Skills in Europe and Beyond* (2024), mckinsey.com/mgi. הדוח מצביע על 12 מיליון מעברי תעסוקה נדרשים באירופה עד 2030, מהלך שמהווה הכפלה של קצב המעברים בעשור שלפני מגפת הקורונה (COVID-19, 2020 - 2022).

32 House of Commons, National Audit Office (UK), *Use of Artificial Intelligence*, לעיל ה"ש 8. הסקר נערך בקרב 87 גופי ממשלה בבריטניה. ראו גם Committee of Public Accounts, *Use of AI in Government*, לעיל ה"ש 8.

33 National Audit Office (UK), *Use of Artificial Intelligence*, לעיל ה"ש 8.

34 U.S. Federal, *AI Use Case Inventory*, לעיל ה"ש 9. נתוני השוואה בין דיווחי 2023 ל-2024 מבוססים על הרישום הפדרלי הציבורי של יישומי AI בסוכנויות. ראו גם U.S. Government Accountability Office (GAO), *AI Use Case Inventory*, לעיל ה"ש 10.

35 Anthropic, *Labor Market Impacts of AI*, לעיל ה"ש 6. הנתון מבוסס על ניתוח שימוש בפועל (observed exposure) ולא על הערכה תיאורטית בלבד. המדד "חשיפה נצפית" מודד במפורש מה AI עושה בפועל ולא מה הוא יכול לעשות.

במסגרת "שותפויות כישרוים" (Skill Partnerships), כפי שמגדיר McKinsey: התכלית אינה ייתור העובדים אלא שינוי תמהיל המיומנויות, כך שלעובדים יש תמריץ אמיתי להכשיר עצמם לתפקידים מועצמי. AI עם גופים אלה נמנים יחידות ביקורת פנים וביקורת המדינה בשירות הציבורי בחלק ממדינות המערב (בין היתר בבריטניה, ארצות הברית ואוסטרליה, כמתועד בדוחות ANAO ו-NAO, GAO המצוטטים בהמשך). יחידות אלה הן שמרניות מטבען ולכן זהירות מאוד במעבר לשיטות עבודה חדשות הדורשות השקעה משמעותית בהדרכה ושינויי פרדיגמות ברמה הארגונית.

הפער בין הפוטנציאל התיאורטי של ה-AI לבין שיעור האימוץ בפועל מתועד היטב בנתוני דוחות שפרסמה חברת לינקדאין⁴⁰. בדוח "Work Change Report: AI Is Coming to Work"⁴¹, שפרסמה החברה בינואר 2025, נמצא כי 88% מהבכירים הגלובליים רואים בהאצת אימוץ ה-AI יעד מרכזי. אולם, דוח שוק העבודה של החברה מינואר 2026, "Building a Future of Work That Works"⁴², מראה שרק כ-3% מחברי הרשת בארצות הברית מציגים כישורי AI בפרופיל שלהם, והם מרוכזים בעיקר במקצועות ההנדסה, ניהול המוצר והמחקר. כלומר, הציפייה הניהולית מתקדמת מהר יותר מהיכולת הארגונית לספק אותה - וזהו הקושי האמיתי שניצב בפני יחידות ביקורת המבקשות להטמיע את הטכנולוגיה. החזון הניהולי לאימוץ AI מתקדם יותר מהמבנה הארגוני הנדרש כדי לתמוך בו.

כל אלה הם אתגרי עומק שאינם טכנולוגיים ונוגעים לתרבות ארגונית, לניהול סיכונים, לנשיאה באחריות עמוקה לגורל הארגון ולהטיות קוגניטיביות של פרטים ושל קבוצות

אם בתדירות קטנה יותר מבני אדם, הרי שגורמי המקצוע שונאי הסיכון יהססו לשאת באחריות למימוש הסיכון. זהו פרדוקס מוכר: החברה מגלה סובלנות רבה יותר לטעויות אנוש מאשר לטעויות מכונה - תופעה הידועה בספרות המחקר כ"סלידה אלגוריתמית" (algorithm aversion)³⁶. הדוגמה של הרכב האוטונומי ממחישה זאת היטב: הציבור דורש מרכבים אוטונומיים סטנדרט בטיחות חמור בהרבה מזה הנדרש מנהגים אנושיים, גם אם הטכנולוגיה מפחיתה סיכונים ביחס לנהיגה אנושית. תאונה שגרמה לה מכונה זוכה לתשומת לב ציבורית ולתגובה פסיכולוגית בעוצמה גבוהה בהרבה מתאונה זהה במאפייניה שגרם לה נהג אנושי³⁷.

ב. אי-ודאות רגולטורית - בסקר Deloitte לרבעון 4 של 2024, שכלל 2,773 מנהלים בכירים מ-14 מדינות, בלט הציות הרגולטורי כחסם המוביל העומד בפני הטמעת טכנולוגיה - שיעור המנהלים שציינו אותו כחסם עלה מ-28% ברבעון 1 ל-38% ברבעון 3³⁸. החסם הזה מורגש במיוחד במוסדות הביקורת הציבוריים: GAO האמריקאי, NAO הבריטי ו-ANAO האוסטרלי - כולם פועלים בסביבה רגולטורית של AI שעודנה מתעצבת, ולכן נוקטים זהירות יתרה בהטמעת AI. חסם רגולטורי נוסף ונפרד מתעצב סביב שאלות של פרטיות, אתיקה, הטעיות ודיפ-פייק - חסם המתגבש במקביל לכניסת EU AI Act לתוקף ולפרסום NIST AI Risk Management Framework³⁹. כלומר, לרשויות ממשלתיות יש סיבה טובה לנהוג בזהירות יתרה.

ג. פערי כישרים וידע - המנהלים שנסקרו בסקר Deloitte מדווחים על פערים בידע וביכולות בתחום ה-AI המונעים הטמעת כלים מתקדמים בארגונים. אתגר זה נפתר רק

Berkeley J. Dietvorst, Joseph P. Simmons & Cade Massey, *Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them* 36 *Err.* 144(1) *J. of Experimental Psych. General* 114 (2015).

Azim Shariff, Jean-François Bonnefon & Iyad Rahwan, *Psychological Roadblocks to the Adoption of Self-Driving Vehicles*, 1 (10) *Nature* 37 (2017).
Azim Shariff, Jean-François Bonnefon & Iyad Rahwan, *How Safe Is Safe Enough?* *Psychological Human Behaviour* 694 (2017).
Mechanisms Underlying Extreme Safety Demands for Self-Driving Cars, 126 *Transportation Rsch. Part C* (2021), art. 103069.

Deloitte AI Institute, *The State of Generative AI in the Enterprise: Generating a New Future* (Quarter 4 Report 2025) 38
www.deloitte.com/content/dam/assets-zone3/us/en/docs/services/consulting/2024/us-state-of-gen-ai-q4.pdf.
הסקר נערך ביולי ספטמבר 2024 בקרב 2,773 מנהלים בכירים ב-14 מדינות.

Regulation (EU) 2024/1689 Laying Down Harmonised Rules on Artificial Intelligence (AI Act), OJ L of 12 July 2024 39
(entered into force 1 August 2024).
התקנה האירופית לרגולציה של בינה מלאכותית, האוסרת את השימוש ב-AI למטרות דיורג חברתי, בעלת דרישות שקיפות ובקרה אנושית על מערכות בסיכון גבוה.

40 הרשת המקצועית הגדולה בעולם שבה יותר מ-1.3 מיליארד חברים. הרשת מפעילה מכון מחקר (Economic Graph Research Institute) המדווח על מגמות בשוק העבודה.

LinkedIn Economic Graph Research Institute, *Work Change Report: AI Is Coming to Work* (2025), economicgraph.linkedin.com 41

LinkedIn Economic Graph Research Institute, *Labor Market Report: Building a Future of Work That Works* (2026) 42
economicgraph.linkedin.com



בארגונים ובמערכות שטרם בשלו להטמעת שינויים משמעותיים בקרבם^{43, 44}.

ממצא נוסף מהמחקר של Anthropic ממרץ 2026 ראו לתשומת לב מיוחדת: עדיין אין עלייה מובהקת באבטלה בקרב עובדים בתחומים חשופים (תחומים שבהם יותר מ-80% מהמשימות ניתנות לביצוע תיאורטי על ידי AI לפי המדד של Anthropic - בראשם מקצועות מחשב, מזכירות, עסקים ומשפט), אך ניכרת האטה מובהקת בגיוס של צעירים בני 22 - 25 לתחומים אלה⁴⁵. על פי דוח LinkedIn Labor Market Report מינואר 2026, הגיוס למשרות מפתחים ירד ב-20% - 35% לעומת תקופת טרום הקורונה, בעיקר בשל חוסר ודאות כלכלית לצד שינויים בתמהיל הכישורים הנדרש⁴⁶. כלומר, השוק כבר מגיב וחווה שינוי, אם כי בצורה שקטה: לא פיטורים המוניים, אלא צמצום הדרגתי של גיוס עובדים חדשים לתחומים בעלי חשיפה גבוהה ל-AI. גם מחקר מעבדת Stanford Digital Economy מצביע על ירידה של 13% - 20% בתפקידי entry-level בתחומי הנדסת התוכנה ושירות הלקוחות.

הכפלה כל 89 יום

ארגון METR (Model Evaluation and Threat Research)⁴⁷, שעוקב באופן שיטתי אחרי יכולתם של סוכני AI לבצע משימות מורכבות באופן אוטונומי, מודד זאת באופן שיטתי ומעדכן את "אופק הזמן" (Time Horizon) - כלומר קובע מהו משך המשימה האנושית שסוכן AI מצליח לבצע באופן עצמאי בהצלחה של 50% לפחות. לפי המדידה העדכנית של הארגון (Time Horizon 1.1, ינואר 2026), זמן ההכפלה משנת 2023 ואילך הוא כ-131 יום (לעומת 165 יום תחת המתודולוגיה הקודמת), והוא ממשיך להתקצר - בנתוני 2024 ואילך הוא עומד על כ-89 יום בלבד⁴⁸. כלומר, לא רק שהיכולות גדלות - הן גדלות בקצב שהולך ומואץ. לעיתים קרובות תחזיות בעניין קצב התפתחות הטכנולוגיה מתיישנות בתוך חודשים בודדים.

בשנת 2020 היו מערכות AI מסוגלות לבצע משימות אנושיות שארכו כמה שניות, כגון סיווג פריט אחד, תיקון שגיאת הקלדה או תשובה למשימת שפה קצרה; בתחילת שנת 2023 הן הצליחו במשימות הדורשות מאדם כמה דקות (כגון סיכום מסמך קצר, ניסוח פסקה או פתרון בעיית תכנות בסיסית); בסוף שנת 2024 הן הצליחו במשימות הדורשות מאדם כמה עשרות דקות (כגון סקירת מסמך ארוך, ניתוח טבלת נתונים מורכבת או כתיבת טיוטת סעיף דוח). בהתאם לקצב המואץ של ביצוע המשימות כמתואר, סוכני בינה מלאכותית עתידים לבצע משימות הדורשות שעה מלאה בעוד כחצי שנה, יום עבודה מלא - בעוד כשנה ושבוע עבודה - בעוד כשנתיים.

על פי שלושת המחקרים האמורים, הטכנולוגיה בת זמננו כבר מסוגלת לבצע למעלה מ-28% מהמשימות הכרוכות בעבודת הביקורת באופן עצמאי לחלוטין, ולכן הפער בין היכולת הטכנולוגית ליכולת הביצועית הולך ומצטמצם בקצב מואץ. דהיינו, בעוד שלוש עד חמש שנים תוכל הבינה המלאכותית לבצע את עבודת הביקורת בשלמותה, בתנאי שכל הגופים המבוקרים יטמיעו במערכותיהם כלי בינה מלאכותית מתקדמים בממשק מאובטח עם המערכות שבשימוש משרד מבקר המדינה ונציב תלונות הציבור.

תרחישי שימוש: AI בעבודת הביקורת בפועל

הדיון על יכולות ה-AI אינו תיאורטי בלבד. הן מוסדות ביקורת עליונים ברחבי העולם (Supreme Audit Institutions - SAls), המאוגדים תחת ארגון INTOSAI הבין-לאומי המונה 193 חברים כדוגמת ה-GAO בארצות הברית וה-NAO בבריטניה, והן יחידות ביקורת פנימית כבר מפתחים ומיישמים כלים מבוססי בינה מלאכותית. להלן חמישה תרחישי שימוש מרכזיים, כל אחד מלווה בדוגמה ממשית מגוף ביקורת שכבר התחיל בשלב היישום.

Jin Li, Feng Zhu & Pascal Hua, *Overcoming the Organizational Barriers to AI Adoption*, Harvard Bus. Rev. (2025), 43 <https://hbr.org/2025/11/overcoming-the-organizational-barriers-to-ai-adoption>; Cole Stryker, Amanda Downie & Matthew Finio (42%), הצדקה עסקית חלשה (42%) וחששות לפרטיות או לסודיות (40%).

5 *Biggest AI Adoption Challenges for 2026*, IBM (n.d.), <http://www.ibm.com/think/insights/ai-adoption-challenges>. 44 לפי סטרייקר, דאוני ופיניו, חמשת החסמים הגדולים לאימוץ AI ב-2026 הם: חשש מדיוק או הטיה (45%), מחסור בנתונים קנייניים (42%), חוסר במומחיות Gen AI.

Anthropic, *Labor Market Impacts of AI*, לעיל ה"ש 6. ההאטה בגיוס נצפתה בקרב עובדים בני 22 - 25 בתחומים עם חשיפה גבוהה ל-AI.

LinkedIn Economic Graph, *Labor Market Report* 46 לעיל ה"ש 44. הדוח קובע במפורש: "hiring remains 20-35% below pre-pandemic levels in advanced economies... AI isn't driving the hiring slowdown".

47 מכון מחקר עצמאי המודד שיטתית את יכולות מודלי AI.

METR, *Time Horizon 1.1: Task-Completion Time Horizons of Frontier AI Models* (2026), <https://metr.org/blog/2026-1-29-time-horizon-1-1/>. 48

ניתוח מסמכים רחב-היקף

אחד האתגרים הגדולים בביקורת הוא היקף החומרים: אלפי חוזים, פרוטוקולים, מכתמים ונהלים שיש לבחון. המבקרים האנושיים קוראים חלק מהחומרים האלה - אולם בעידן ה-AI, החומרים רבים כל כך שגם דגימה סטטיסטית נכונה עלולה ליצור הטיה בממצאים. משכך, לעיתים קשה לדעת אם אכן נבחן מדגם מייצג.

ה-GAO האמריקאי (Government Accountability Office) מפעיל מערכת המבוססת על מודל שפה לסינתזה של דוחות ביקורת קודמים, סיוע בסקירות עריכתיות וסריקת מסמכי קונגרס⁴⁹. פעילות שדרשה בעבר הקצאת משאבי זמן וארכה שבועות ארוכים של קריאה, ניתוח ועיבוד של מידע כתוב נעשית כעת בתוך שעות, עם כיסוי מלא של כל החומר הרלוונטי. סקירת הבינה המלאכותית מבססת את היכולת של הביקורת לבנות עבור הציבור תמונת מצב עדכנית ומדויקת של המציאות הנבדקת בגופים המבוקרים.

זיהוי חריגות ודגלים אדומים

ביקורת פיננסית שמרנית כוללת בדיקה של היקף עסקאות מסויים. הערך המוסף של פעילות הבינה המלאכותית הוא בסריקת כל הנתונים ללא מגבלה וזיהוי דפוסים וחריגות שבינה אנושית נוטה להתעלם מהם או שהייתה מתקשה לזהות בשל ההיקף העצום של הנתונים. יכולות אלה תועדו במחקר אקדמי על data-driven auditing ובמסגרת מסמך הגדרת האחראיות של GAO ל-AI במגזר הציבורי⁵⁰.

Australian National Audit Office - מוסד הביקורת העליון של אוסטרליה (ANAO) בחן כיצד רשות המיסים (ATO) משתמשת ב-AI לזיהוי הונאות ומצא שהכלים מזהים חריגות שנשמטו ממבקרים אנושיים בשל מגבלות הזמן והיקף הנתונים. סוכן AI שמנתח "כל עסקה" לא רק עושה זאת במהירות רבה יותר - הוא מבטל לחלוטין את מגבלת המשאבים, כלומר שעות העבודה והתקציב של מוסד

הביקורת, שבהכרח מוגבלים יחסית להיקף הפעילות המבוקרת⁵¹.

ניסוח ממצאים והמלצות

פרסום ממצאי הביקורת לציבור מצריך ניסוח ברור של ממצאים בלשון ברורה המשקפת בסיס עובדתי מוצק, התאמה לנומרה והמלצה לתיקון. המדריך לביקורת המדינה מורה כי "ממצא יתואר באופן אובייקטיבי ותמציתי ללא פרשנות של צוות הביקורת או הצגת עמדתו בנושא"; כי יש לתאר אותו "כהשוואה בין המצב הרצוי למצב המצוי" ולנסחו כך ש"יעמוד בפני עצמו"; וכי בתיאורו "ישמשו המילים 'בביקורת עלה כי...' או 'נמצא כי...'". תיאור ממצא הוא, אפוא, משימת כתיבה שדורשת חדות, בהירות ועמידה בכללים מתודולוגיים מדויקים⁵².

מודלי שפה גדולים מסוגלים לעבד נתונים גולמיים ולנסח על בסיסם ממצאים מובנים - בהתאם לפורמט, לסגנון ולרמת הפירוט של דוחות ביקורת קיימים. כלי בינה מלאכותית כגון Thomson Reuters CoCounsel Audit ו-Wolters Kluwer CCH Xcess Expert AI⁵³ מסוגלים אף ליצור טיוטה ראשונית של ממצא, לנמק אותו ביחס לסרגל הביקורת, להציע המלצות ואף להצליב את הנתונים עם ממצאים דומים מביקורות קודמות. כך נוצרת סינתזה בין דוחות ביקורת מזמנים שונים המעשירה את הממצאים ומעניקה ערך מוסף לציבור. הבינה המלאכותית יכולה לסייע גם בניסוח ובתרגום לשפות שונות, שהם אתגר מובהק בישראל לאור חובת הפרסום בעברית, ערבית ואנגלית.

הערכת סיכונים ותיעודם

תכנון ביקורת מצריך הערכה מדורגת של סיכונים: באילו תחומים עלולים להתרחש כישלונות, אסונות, תאונות, או הפרות חוק משמעותיות. המבקר צריך למפות את סדרי העדיפויות בתוך מנעד אפשרויות, לשקול את השיקולים

49 Artificial Intelligence (GAO), U.S. Government Accountability Office, לעיל ה"ש 18. GAO מתאר במסמך את פריסת מודל השפה הגדול ככלי לסינתזה של דוחות עבר, סיוע בסקירות עריכתיות וסריקת מסמכי קונגרס.

Deniz Appelbaum, Alexander Kogan & Miklos A. Vasarhelyi, *Big Data and Analytics in the Modern Audit Engagement*: 50 U.S. Government Accountability Office *Research Needs*, 36(4) Auditing: A Journal of Practice & Theory 1 (2017); (GAO), Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities (GAO-21-519SP) (2021), <https://www.gao.gov/products/gao-21-519sp>.

Australian National Audit Office (ANAO), Governance of Artificial Intelligence at the Australian Taxation Office, Auditor-General Report 51 (No. 36 of 2023-24) (2025), www.anao.gov.au/work/performance-audit/governance-of-artificial-intelligence-the-australian-taxation-office.

52 משרד מבקר המדינה ונציב תלונות הציבור, *מדריך לביקורת המדינה*, לעיל ה"ש 26.

Thomson Reuters, *CoCounsel Audit*, thomsonreuters.com/en/products/cocounsel.html; Wolters Kluwer, *CCH Xcess Expert AI*, 53 wolterskluwer.com/en/solutions/cch-xcess/expert-ai.

ביקורת על גופים מבוקרים המשתמשים בבינה מלאכותית

ככל שמערכות הבינה המלאכותית הולכות וחודרות לממשל ולמגזר הציבורי - מקבלות החלטות על זכויות, מנהלות מכרזים, מזהות הונאות ומספקות שירותים לתושבים - מוסדות הביקורת נדרשים לפתח יותר ויותר יכולות חדשות לבדיקת מערכות אלה.

הקופסה השחורה

מערכות AI - ובפרט מודלים עמוקים ומודלי שפה גדולים - מתוארות לעיתים קרובות כ"קופסה שחורה": הן מקבלות קלט ומפיקות פלט, אך התהליך הפנימי אטום ולעיתים גם מפתחיו אינם מסוגלים להסבירו. עבור מבקר זהו אתגר מהותי. כיצד מבקרים מערכת שאפילו מפתחיה אינם מסוגלים תמיד להסביר את החלטותיה? כיצד מעריכים אם מערכת AI שמשמשת לקבלת החלטות בנוגע לתושבים - זכאות לגמלה, דירוג סיכון אשראי, אכיפה ברחבי וכו' - פועלת בצורה הוגנת, שוויונית ותקינה?

דוגמאות בין-לאומיות

ה-ANAO האוסטרלי ביצע ביקורת ביצועים על שימוש רשות המיסים (ATO) ב-AI ומצא כי לרשות לא הייתה מסגרת ממשל AI מספקת וכי פעלה ללא סיווג סיכונים ברור וללא מנגנוני בקרה הולמים.⁵⁸

ה-Bundesrechnungshof הגרמני הלך צעד נוסף קדימה ופיתח מתודולוגיה שלמה לביקורת מערכות אלגוריתמיות, כולל בדיקות שקיפות, הוגנות, אחריותיות ובטיחות. המתודולוגיה הזו היא אבן דרך חשובה עבור מוסדות ביקורת אחרים שמחפשים מסגרת עבודה לנושא שעדיין בחיתוליו.⁵⁹

ה-NAO הבריטי סקר את אופן הפריסה של AI ברשויות

והאיזונים הרלוונטיים בין הצרכים המשתנים על פי האינטרס הציבורי ולהחליט באילו נושאים על משרד להתמקד בתוכניות העבודה השנתיות והרב-שנתיות.

ה-NAO הבריטי דיווח על שימוש בכלי AI לטובת ניתוח נתוני המגזר הציבורי לצורך תיעוד נושאי הביקורת. במקום להסתמך על שיקול דעת בלבד - שמושפע מזמנות מידע, מהטיות קוגניטיביות ומנטייה לחזור לנושאים מוכרים - צוותי הביקורת משלבים ניתוח אלגוריתמי שמצביע על תחומי הסיכון הגבוהים ביותר ומאפשר הקצאת משאבים רציונלית ומבוססת נתונים.⁵⁴

סקירת רגולציה והשוואה בין-לאומית

לשם קיום ביקורת השוואתית ויצירת תשתית לגיבוש ממצאי ביקורת עדכניים, המבקר נדרש במסגרת עבודתו השגרתית להיות בקיא בחקיקה וברגולציה הקיימת הן במדינות מתקדמות והן בארגונים בין-לאומיים. מדובר בכמות בלתי רגילה של חומרים בשפות שונות, אשר מומחיות בתוכניהם מצריכה השקעת משאבים ניכרים. לעומת זאת, שימוש בכלי בינה מלאכותית לצרכים האמורים עשוי לחסוך זמן וממון רבים ולהניב תוצרים יעילים ומועילים.

בשנים האחרונות מתגבשת קהילה בין-לאומית סביב השימוש במערכות בינה מלאכותית לצורכי ביקורת, כפי שעולה משורת יוזמות של מוסדות ביקורת עליונים. ה-NAO הבריטי פרסם שני מסמכי הנחיה: 'How to audit 'Good practice guide-I artificial intelligence models' for organisations using AI';⁵⁵ וה-GAO האמריקאי משתמש במודל שפה פנימי לסיכום דוחות קודמים, סיוע בסקירות עריכתיות וסריקת מסמכי קונגרס;⁵⁶ וה-ANAO האוסטרלי ערך ב-2023 ביקורת על ממשל AI ברשות המיסים (ATO) שכללה שבע המלצות שכולן התקבלו.⁵⁷ במקביל, כלי AI מאפשרים למוסדות הביקורת עצמם לעקוב בזמן אמת אחר שינויי רגולציה בעולם, לבצע השוואות בין-לאומיות ולזהות פערים בין הנורמה המקומית לבין שיטות מיטביות.

54 National Audit Office (UK), Use of Artificial Intelligence, לעיל ה"ש 8.

55 Daniel Lambauer, How to Audit Artificial Intelligence Models, NAO Insight (2021), www.nao.org.uk/insights/how-to-audit-artificial-intelligence-models; National Audit Office (UK), Good Practice Guide for Organisations Using AI (2026), www.nao.org.uk/insights/good-practice-guide-for-organisations-using-ai.

56 U.S. Government Accountability Office, Artificial Intelligence, לעיל ה"ש 18.

57 Australian National Audit Office (ANAO), Auditor-General Report, לעיל ה"ש 52.

58 Governance of AI at ATO, הביקורת מצאה כי ל-ATO לא הייתה מסגרת ממשל AI מספקת.

59 SAs of Brazil, Finland, Germany, the Netherlands, Norway and the UK (incl. Bundesrechnungshof), Auditing Machine Learning Algorithms - A White Paper for Public Auditors (2020 updated 2023, 2025), auditingalgorithms.net

המבקר עדיין מקבל את כל ההחלטות, אך הכלי מבצע חלק ניכר מהעבודה הפרוצדורלית. האנלוגיה המתאימה היא נהיגה עם מערכת עזר: הנהג אחראי, ידיו על ההגה, אך המערכת עושה את רוב מלאכת הנהיגה.

רמה 3: סוכן אוטונומי (Agent) - מבקר דיגיטלי

זהו השדרוג המשמעותי ביותר של כלי הבינה המלאכותית בעת הזאת. סוכן AI שמקבל הנחיה כללית - "בצע ביקורת ציות על מחלקת הרכש ביחס לתקנות המכרזים" - ומבצע את כל התהליך באופן עצמאי: ניגש למערכות המידע, שואב נתונים ומסמכים, מנתח אותם ביחס לנוכחות, מזהה פערים וחריגות, מנסח ממצאים ומפיק דוח. המבקר האנושי הופך ממבצע למפקח - קורא את הדוח, מאמת ממצאים מרכזיים ומפעיל שיקול דעת בסוגיות רגישות.

סוכן AI שיוצר קשר עם מערכות הגוף המבוקר, שואב חומרים, מנתח פערים, מגבש ממצאים ומפיק דוח - זהו תרחיש שהטכנולוגיה כבר מאפשרת היום, ובכל כמה חודשים הוא הופך לבשל ואמין יותר. השאלה המעשית אינה אם הדבר אפשרי, אלא מתי הוא יגיע לרמת אמינות שמאפשרת הסתמכות עליו, ומה תידרש המסגרת המבוקרת לאפשר.

התהליך מול התוצר: האינטרס הציבורי בקיום ביקורת אנושית

האם לציבור יש אינטרס בכך שעבודת הביקורת תיעשה דווקא בידי אדם? בהקשר זה, יש להבחין בין שני ערכים שונים שהביקורת נועדה לשרת:

הערך שבתהליך מול הערך שבתוצר

מאז הקמתם של מוסדות ביקורת המדינה נתפסה עבודת הביקורת כתהליך אנושי מיסודו. המבקר אינו רק מנתח נתונים - הוא נוכח פיזית, מנהל שיחות, שואל שאלות שלא ניתן לנסח מראש, מזהה אווירה ותרבות ארגונית ומפעיל שיקול דעת מקצועי שמבוסס על ניסיון מצטבר⁶³. הנוכחות האנושית עצמה - עצם הידיעה שהארגון חשוף

הממשל ומצא שרק 30% מהגופים שהטמיעו בעבודתם AI דיווחו על מסגרת ממשל ייעודית. זהו ממצא מדאיג שמצביע על פער ניכר בין קצב האימוץ הנלהב של הטכנולוגיה לבין רמת הבקרה על יישומיה⁶⁰.

חוק ה-AI של האיחוד האירופי (EU AI Act), שנכנס לתוקף ב-2024, מחייב לסווג מערכות AI לפי רמות סיכון ומטיל חובות שקיפות, תיעוד ובקרה על מערכות בסיכון גבוה. גופי הביקורת בעולם יידרשו לציית לדרישות אלה ולפתח מומחיות טכנולוגית שתחייב שדרוג משמעותי של יכולות הצוותים⁶¹. בין היתר יחויבו הצוותים להקפיד על עקרון human-in-command (HIC) - אחת משלוש רמות הפיקוח האנושי שהגדיר ה-High-Level Expert Group on-AI של נציבות האיחוד האירופי, לצד human-in-the-loop (HITL) ו-human-on-the-loop (HOTL)⁶². כלומר, גם כאשר הביקורת נעזרת בכלי AI, שיקול הדעת המקצועי והאחריות לניסוח הסופי נשארים בידי המבקר האנושי.

מצ'ט לסוכן: רמות האוטונומיה של הכלי

לא כל השימושים בבינה מלאכותית שווים. יש ספקטרום רחב בין צ'טבוט פשוט שעונה על שאלות ובין סוכן אוטונומי שמבצע ביקורת מלאה. הבנת הספקטרום הזה חיונית לכל דיון רציני על עתיד הביקורת, שכן כל רמה מציבה דרישות שונות של פיקוח, אמון, תשתית ומסגרת נורמטיבית.

רמה 1: צ'ט - מנוע שאלות ותשובות

בינה מלאכותית ברמה בסיסית זו מתייחסת אך ורק לשאלות המסוימות ששואל המבקר. למשל: "מה קובע חוק חובת המכרזים לגבי פטורים?" או "סכם לי את עשר הנקודות המרכזיות בדוח הזה". זוהי הרמה הנפוצה ביותר היום, והיא כבר חוסכת שעות עבודה רבות של חיפוש ידני. המבקר נשאר בשליטה על התהליך. הכלי משמש מנוע חיפוש מתקדם וידידותי למשתמש.

רמה 2: עוזר (Copilot) - שותף לעבודה

הכלי לא רק עונה אלא גם יוזם: מציע ממצאים, מתריע על חריגות בנתונים ומנסח טיוטות ראשוניות של קטעים בדוח.

National Audit Office (UK), *Use of Artificial Intelligence*, 60 לעיל ה"ש 8. הסקר מצא כי רק 30% מהגופים שמתמשים ב-AI דיווחו על מסגרת ממשל ייעודית.

European Commission, EU AI Act, Regulation 2024/1689, August 2024. 61

European Commission, *Ethics Guidelines for Trustworthy AI*, European Commission (2019), digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai. ההנחיות מבחינות בין שלוש רמות פיקוח אנושי: התערבות אנושית בכל מחזור החלטה - HITL; ניטור פעילות המערכת בזמן ריצה ויכולת התערבות בה - HOTL; ושיקול דעת בבחינת המסגרת הכוללת וההחלטה אם ומתי להשתמש במערכת - HIC.

International Organization of Supreme Audit Institution (INTOSAI), *ISSAI 100* 63, לעיל ה"ש 1. העקרונות מגדשים את חשיבות העצמאות, האובייקטיביות ושיקול הדעת המקצועי.

מלאכותית יכול לבצע את אותה עבודה בתוך ימים או שבועות. יתרון היעילות אינו מתבטא רק בקיצור לוחות זמנים: הוא מאפשר ביקורת רציפה (continuous auditing) במקום ביקורת נקודתית, וכך מאפשר לאתר חריגה בזמן אמת ולא שנים אחריה. בגופי ביקורת מתקדמים כמו ה-GAO האמריקני וה-ANAO האוסטרלי הובהר כי היעילות היא יעד מוביל לאימוץ AI, ולא תוצר לוואי שלו⁶⁵. יתרה מכך, יעילות הביקורת האנושית מושפעת מהסתמכות על מדגמים מצומצמים, וכך מטבעה היא אפקטיבית במידה חלקית בלבד: היא מאתרת ליקויים בגורמים שנדגמו אך מחמיצה לעתים את השיטה הכוללת. בינה מלאכותית מאפשרת סריקה של 100% מהעסקאות, 100% מהמסמכים ו-100% מההחלטות, ובכך מעלה את יעילות הביקורת לרמה שלא הייתה אפשרית עד כה. בד בבד, היא מאפשרת זיהוי מדויק יותר של חריגות, לעומת מדגם אנושי הסובל מ"רעש" הנובע מהטיות אישיות (noise) ומשונות קבוצתית, כפי שהוכיחו כהנמן ועמיתיו⁶⁶.

חיסכון: צמצום עלויות והרחבת ממצאי הביקורת

עלות שעת עבודה של מבקר מקצועי במשרד מבקר המדינה נאמדת במאות שקלים, וכוללת משכורת, הכשרה והוצאות תפעול. עלות שעת עבודה של סוכן AI מתקדם, לאחר הפריסה ראשונית, נאמדת באגורות. אין מדובר רק בחיסכון תקציבי - זהו שינוי מהותי באופן הביקורת האפשרי: ניתן לבקר פי עשרה גופים באותו תקציב, לבחון פי מאה מסמכים ולכסות תחומים שלמים שכיום אינם נבדקים כלל בשל מחסור במשאבים. משמעותו של חיסכון כזה אינה ביצוע פחות פעולות, אלא השגת כיסוי נרחב יותר עם אותה השקעה, וזה התרגום המדויק של הערך הציבורי שבחיסכון.

יעילות וחיסכון אינם ערכים המנוגדים לערכי התהליך והתוצר שנדונו לעיל אלא משלימים אותם. ביקורת שאינה יעילה ואינה חסכונית אינה מממשת את תכליתה הציבורית גם אם היא אנושית לחלוטין. ולהפך, ביקורת ממוכנת שממזה את הערכים הללו אינה מוותרת על ערכים מוסריים אלא ממזה אותם באופן שלם יותר. האתגר אינו להכריע בין הבינה האנושית לבינה המלאכותית, אלא לוודא שהאינטרס הציבורי של יעילות וחיסכון לא ייוותר חסר מימוש בשל היצמדות נוסטלגית לדפוסי עבודה אנושיים בלבד.

באופן תמידי לנכחות של גורם חיצוני לו הבודק לעומק את פעילותו, היא מנגנון הרתעה רב עוצמה.

יתרה מכך, האינטראקציה בין המבקר לגורמים בגוף המבוקר היא חלק בלתי נפרד מתהליך ההשפעה: ההסברים, השאלות, חילוקי הדעות, תהליך ההערות והשיח המקצועי - כל אלה מעצבים לא רק את דוח הביקורת אלא את ההתנהגות של הגוף המבוקר עצמו. המבקר, במובן מסוים, הוא לא רק מפקח אלא גם מחנך. מנקודת מבט זו, החלפת המבקר האנושי במכונה היא לא רק שינוי טכני - היא ויתור על ערך מהותי שהציבור מעוניין בו.

מנגד, ניתן לטעון שבמבחן התוצאה, התוצר הסופי מממש את האינטרס הציבורי המובהק: דוח ביקורת מדויק ומקיף המופק בזמן קצר ומניע את הגופים המבוקרים לתקן ליקויים. אם מכונה מסוגלת להפיק דוח טוב יותר - מדויק יותר, מהיר יותר, עם כיסוי רחב ומלא יותר - מדוע צריך הציבור לשלם עבור תהליך אנושי יקר, איטי וחלקי? לפי גישה זו, לזהות הגורם המבצע את הביקורת אין השפעה מהותית כל עוד התוצר משרת את האינטרס הציבורי ומניע תיקון ליקויים.

האינטרס הציבורי בביקורת ממוכנת: יעילות וחיסכון

הדיון בערך שבתהליך ובערך שבתוצר התמקד בביקורת האנושית, אך הוא חסר טעם ללא הצד האחר של המשוואה: האינטרס הציבורי בביקורת הממומשת באמצעות מכונה. סעיף 2(ב) לחוק יסוד מבקר המדינה קובע את תכלית הביקורת ביחס לגופים המבוקרים: "מבקר המדינה יבחן את חוקיות הפעולות, טוהר המידות, הניהול התקין, היעילות והחסכון של הגופים המבוקרים, וכל ענין אחר שיראה בו צורך"⁶⁴. כאשר מדובר בהטמעת AI בעבודת הביקורת, השאלה המעשית הנגזרת היא אחרת: האם השימוש ב-AI מאפשר למבקר לזהות באופן יעיל ומקיף יותר ליקויים בחוקיות, בטוהר המידות, בניהול התקין, ביעילות ובחיסכון של הגופים המבוקרים. במישור זה הבינה המלאכותית מסוגלת להעצים מאוד את כושר הביקורת.

יעילות: הספק ומהירות כמענה על הפיגור הכרוני

בהסתמכות על ביקורת אנושית יחלפו חודשים ואף שנים מרגע התרחשות המעשה ועד פרסום הדוח. סוכן בינה

64 חוק-יסוד: מבקר המדינה, ס"ח התשמ"ח.

65 U.S. Government Accountability Office (GAO), Artificial Intelligence, לעיל ה'ש 51. המסמך מגדיר ארבעה עקרונות אחריות ל-AI ציבורי: ממשל, נתונים, ביצועים וניטור רציף.

66 Daniel Kahneman, Olivier Sibony & Cass R. Sunstein, Noise: A Flaw in Human Judgment, Little, Brown Spark (2021). 66 שיפוט אנושי בתחומי הרפואה, המשפט והביקורת סובל מ"רעש" (שונות לא רצויה בין מחליטים) ברמה שמסכנת את איכות ההחלטות שיתקבלו.

האם הבינה המלאכותית תייתר את עובדי הביקורת?

ההערכה המבוססת על הטכנולוגיה הקיימת היא שעד שנת 2030, וככל שיותר גופים מבוקרים יטמיעו במערכותיהם כלי בינה מלאכותית, יפחת הצורך בבינה אנושית לשם ביצוע עבודת הביקורת. הנתונים שהוצגו בפרקים הקודמים מדברים בעד עצמם: 57% משעות העבודה כבר ניתנות לאוטומציה, הפער בין הפוטנציאל (80% - 94% בתחומים רלוונטיים) לשימוש בפועל (15% - 33%) נובע בעיקר מהקצב האנושי ולא מהיכולת הטכנולוגית, ולפי מדידת METR העדכנית מינואר 2026 (Time Horizon 1.1) יכולותיהם של סוכני AI מוכפלות בקצב של כ-131 יום על פי כלל הנתונים משנת 2023 ואילך, וכ-89 יום בלבד על פי נתוני 2024 ואילך - קצב שמעיד כי ביצוע משימות של יום עבודה מלא יהיה בהישג יד בטווח הקרוב.

אולם חשוב להבחין בין שתי רמות פעילות של כלי הבינה המלאכותית: הרמה הטכנית והרמה החברתית-כלכלית. מבחינה טכנית, יש לבינה המלאכותית יכולת לבצע ביקורת בהיקף ובאיכות ההולמים את זו של הבינה האנושית. ברמה החברתית-כלכלית - מדובר באיגמה. כיצד ארגונים ינהלו מעבר תפעולי כה דרסטי? כיצד יתפרנסו עובדים שתפקידם יתייתר? כיצד שומרים על הידע האנושי המצטבר בתקופת המעבר? אלה שאלות חשובות ודחופות - אך הן לא צריכות לגרום לנו להתכחש לתשובה על השאלה הראשונה.

דפוס ההשפעה שכבר נצפה בשוק העבודה מחזק את ההערכה. אין עדיין פיטורים המוניים, אך יש האטה מובהקת בגיוס עובדים צעירים לתחומים עם חשיפה גבוהה ל-AI.⁶⁷ השוק מגיב בדרך שקטה אך משמעותית: לא בפיטורים המוניים, אלא בצמצום הדרגתי של גיוס העובדים החדשים. מי שעובד בביקורת היום אינו צריך לחשוש - אבל הוא צריך להיערך. על מי שמתכנן קריירה בתחום הביקורת לדעת שהמקצוע שאליו ייכנס ייראה שונה מהותית מהמקצוע המוכר לו כיום.

ראוי לציין שהפורום הכלכלי העולמי כבר זיהה את השליטה ב-AI כמיומנות הצומחת ביותר בשוק העבודה, עם הכפלה פי שבעה בשנתיים.⁶⁸

הגורמים המשפיעים על קביעת האינטרס הציבורי

ההכרעה בין שתי הגישות המתוארות לעיל בנוגע לקביעת האינטרס הציבורי אינה בינארית, והיא תלויה בכמה גורמים משפיעים.

הגורם הראשון נוגע לקצב האימוץ וההטמעה של מערכות AI בגופים המבוקרים עצמם. ככל שגופים מבוקרים עוברים לעבודה מבוססת AI - קבלת החלטות אלגוריתמית, מערכות אוטומטיות, שירותים דיגיטליים וכד' - כך נדרשת גם הביקורת עליהם להתאים את כליה. אם הגוף המבוקר הוא מכונה, אולי גם מי שמבקר אותו צריך להיות מכונה. ביקורת אנושית על מערכת אלגוריתמית שמקבלת אלפי החלטות בשנייה תהיה ככל הנראה חסרת משמעות מעשית.

הגורם השני תלוי באפיון הגוף המבוקר. יש גופים מבוקרים שבהם הגורם האנושי בביקורת עשוי להישמר גם בעידן ה-AI בשל אופיים החשאי, ובפרט מערכת הביטחון, גופי מודיעין או מערכות ליבה מדינתיות אחרות. בגופים אלה שיקול הדעת האנושי, הצורך במידור, הרגישות של המידע והמורכבות הפוליטית עשויים לדרוש נוכחות אנושית לצד כלי AI, ולא במקומם.

בגופי הביקורת על מערכות אלה יהיה ככל הנראה מקום להמשיך ולהעסיק מבקרים אנושיים - אך תפקידם ישתנה מהותית. משינוי תהליך העבודה (מכותב לעורך), דרך פיתוח מיומנויות חדשות (הנדסת פרומפטים, פיקוח על סוכנים, בדיקת מודלים ועוד) ועד להגדרת תפקידים היברידיים כגון "מבקר-טכנולוג" - ראו הרחבה בפרק המתווה האופרטיבי שלהלן.

טענת המאמר אפוא היא כי כאשר הביקורת האנושית הקיימת מכסה שבריר מהמציאות הציבורית, מגיעה באיחור ונסמכת על נתונים חלקיים, אזי הדרישה להישען על בינה אנושית גרידא אינה מובנת מאליה - היא עצמה דורשת הצדקה. הנטל אינו מוטל עוד על מי שמציע להחליף את המבקר האנושי, אלא על מי שטוען שאין להמירו.

מן הדיון הנורמטיבי, נעבור עתה לבחינה כמותית של אופק הזמן: מתי, על בסיס נתוני שוק העבודה והקצב המואץ של יכולות הבינה המלאכותית, עשויה ההחלפה להתממש בפועל.

67. LinkedIn Economic Graph Research Institute, *Labor Market Report* 67, לעיל ה"ש 44. הדוח מתעד צמצום בגיוס עובדים צעירים בתחומים חשופי AI.

68. World Economic Forum, *Future of Jobs Report* (2025), https://reports.weforum.org/docs/WEF_Future_of_Jobs_Report_2025.pdf. 68 הדוח מצביע על "AI fluency" כמיומנות הצומחת ביותר בשוק העבודה.

סיכונים ואתגרים

אין לזלזל בסיכונים שבשילוב בינה מלאכותית בעבודת ביקורת. דווקא מי שמכיר את היכולות המרשימות בתחום צריך להכיר גם את המגבלות הממשיות - ולנהל אותן באחריות.

הזיות - Hallucinations

מודלי שפה גדולים "ממציאים" לעיתים מידע - מצטטים מקורות שאינם קיימים, מציגים נתונים שגויים או מגבשים ממצאים שאין להם בסיס עובדתי. בתחום הביקורת, שבו דיוק עובדתי הוא היסוד לכל ממצא והמלצה, זהו סיכון חמור שאין להקל בו ראש. כל פלט של AI חייב לעבור אימות - אנושי או אוטומטי - ואין לקבלו כעובדה ללא בדיקה צולבת.

הטיות - Bias

מודלי AI נבנים על נתוני אימון שמשקפים הטיות קיימות בחברה ובארגונים. אם מודל אומן על דוחות ביקורת היסטוריים, הוא עלול לשחזר הטיות מוכרות - למשל מיקוד יתר בגופים מסוימים על חשבון אחרים, העדפת סוגי ממצאים מסוימים או התעלמות שיטתית מתחומים שלא קיבלו תשומת לב בעבר. מבקר שמסתמך על AI ללא מודעות להטיות אלה עלול, שלא במתכוון, להנציח בעיות במקום לחשוף אותן.

אבטחת מידע וסודיות

חומרי ביקורת כוללים לעיתים קרובות מידע רגיש ביותר - נתונים כספיים של גופים מבוקרים, מידע אישי של נבדקים, מידע מסווג, נתונים חסויים או כאלה שיש להותירם אנונימיים. השימוש בכלי AI חיצוניים מעלה שאלות מהותיות על הריבונות על מידע: היכן מעובד המידע, מי נחשף אליו, האם הוא משמש לאימון מודלים ועוד. מוסדות ביקורת יצטרכו לפתח מדיניות ברורה ומחייבת ובמקרים רבים להפעיל מודלים פנימיים בסביבה סגורה ומאובטחת.

אתיקה ואחריותיות

מי אחראי לממצא ביקורת שנוצר באמצעות AI? שאלת "חתימת המבקר" אינה פורמלית בלבד. יש לה השלכות משפטיות, מקצועיות ואתיות כבדות משקל. אם סוכן AI מפיק ממצא שגוי שמוביל לנזק לגוף מבוקר או לפגיעה באמון הציבור - מי נושא באחריות? המבקר שהפעיל את הכלי? מנהל יחידת הביקורת? ספק הטכנולוגיה? שאלות אלה דורשות מענה נורמטיבי מוסדר שעדיין אינו קיים ברוב המדינות.

תלות טכנולוגית

מוסד ביקורת שמסתמך על מערכות AI של ספקים חיצוניים עלול למצוא עצמו תלוי בטכנולוגיה, בעדכונים,

בתמחור ובזמינות. שאלת הריבונות הדיגיטלית רלוונטית במיוחד למוסדות ביקורת מדינה, שעצמאותם ואי-תלותם הם ערכי יסוד. בניית יכולות פנימיות, הכשרת צוותים טכנולוגיים ושמירה על אופציות מגוונות יכולים להיות חלק מהפתרון לאתגרים הללו.

ממתודולוגיה ליישום: מפת דרכים

ההכרה ביכולת הטכנולוגית אינה מספיקה - צריך גם דרך פעולה מעשית. להלן הצעה למפת דרכים בשלושה אפיקים, המותאמת הן למוסדות ביקורת מדינה והן ליחידות ביקורת פנימית בארגונים.

בטוח קצר: ההיכרות ושימוש בסיסי

- הכשרת צוותי ביקורת בשימוש בכלי AI קיימים - צ'אטבוטים מתקדמים, כלי סיכום מסמכים, מנועי חיפוש חכמים וכד'. הגדרת מדיניות שימוש ברורה: מה מותר, מה אסור ואילו סוגי מידע אין להעלות לכלים חיצוניים.
 - מינוי והכשרה של "שגריר AI" בכל יחידת ביקורת - גורם שתפקידו לזהות הזדמנויות, להנגיש כלים ולשמש כגשר בין הטכנולוגיה לצוות.
- המטרה המעשית בשלב זה: כל מבקר ישתמש ב-AI בעבודתו השוטפת.

בטוח הבינוני: אינטגרציה והתאמה

- פיתוח או רכישה של כלים ייעודיים לביקורת - מערכות שמחוברות למאגרי המידע הרלוונטיים ומותאמות לתהליכי העבודה הייחודיים של יחידת הביקורת.
- הרצת פיילוט של סוכני AI בנושאי ביקורת ספציפיים, לדוגמה ביקורת ציות, ביקורת רכש או ניתוח תקציבי.
- פיתוח מתודולוגיה לביקורת על מערכות AI בגופים מבוקרים והגדרת מסגרת ממשל AI פנימית.

בטוח הארוך: אוטונומיה ושינוי מבני

- פריסת סוכנים אוטונומיים לביקורת, עם פיקוח אנושי מדורג.
- שילוב מערכתי מלא של AI בתהליכי הביקורת, מתכנן ועד מעקב.
- פיתוח מודלים מותאמים אישית שמאומנים על דוחות ביקורת היסטוריים, חקיקה ונהלים רלוונטיים.

הטמעת בינה מלאכותית בתוך מוסדות הביקורת עצמם

הצעד הראשון הוא הטמעה מלאה של כלי AI בעבודתו השוטפת של מוסד הביקורת: ניתוח מסמכים, סריקת מאגרי נתונים, זיהוי חריגות ואיתור דפוסים. זאת לצד רכש או פיתוח של כלי ביקורת ייעודיים המחוברים למאגרי המידע של הגופים המבוקרים ולשירותי גילוי ראיות אלקטרוני⁷⁰ של מוסד הביקורת. על מוסד הביקורת להיות חלוץ בהטמעתה של הטכנולוגיה שהוא עתיד לבקר.

ביקורת על מערכות בינה מלאכותית בגופים המבוקרים

בד בבד, על מוסד הביקורת לפתח מתודולוגיה ייחודית לביקורת על מערכות AI בגופים מבוקרים. בעידן שבו משרדי ממשלה, רשויות רגולטוריות וחברות ממשלתיות משלבים אלגוריתמים בתהליכי קבלת החלטות - החל מקבלה למוסדות חינוך וכלה בהקצאת טיפול רפואי דחוף - על המבקר לבחון את טיב הנתונים, את הוגנות התוצאות, את קיומה של בקרה אנושית ואת עמידתם של המודלים בהוראות הדין ובכללי האתיקה. מסגרת ה-GAO למבקרי בינה מלאכותית מגדירה את העקרונות הללו במפורש: ממשל, נתונים, ביצועים וניטור רציף.

ממשקים טכנולוגיים בין-ארגוניים

כדי שכלי הבינה המלאכותית המוטמעים במערכות המבקר יוכלו לקיים דיאלוג ישיר עם כלי הבינה המלאכותית שבמערכות הגוף המבוקר, יש לקיים תנאים אחדים: קיום תקני ממשק פתוחים (Open APIs), הסדרה חוזית של הרשאות גישה, יצירת שכבת אימות זהות אלגוריתמית והגדרת פרוטוקולי אבטחה הדדיים. כדי להבטיח עצמאות מקצועית ויכולת ביקורת רציפה שאינה תלויה ברצון הטוב של המבוקר, על מוסד הביקורת לקבוע את הדרישות, ולא רק לצרוך את המידע שהגוף המבוקר מספק לו⁷¹.

פיילוטים מבוקרים

המעבר ממצב של עבודה אנושית לעבודה הנתמכת בבינה מלאכותית מחייב הפעלת פיילוטים מבוקרים. אלה יערכו בצורה מובנית: בחירת נושא ביקורת שמאפיין תכנים דיגיטליים מובנים (כגון רכש, תשלומים או ציוד); הרצת

- התאמת המבנה הארגוני: פחות מבקרים "קלאסיים" ויותר מפקחים על מערכות AI, מנהלי נתונים ומומחי מתודולוגיה. גם תפקידי הפיקוח הללו נכונים לשלב המעבר ולא למתכונת הקבע של המקצוע⁶⁹.

יש לשקול גורם נוסף: בתנאים הישראליים, שבהם מוסדות ציבוריים פועלים עם משאבים מוגבלים ובהיקף קטן יחסית למדינות גדולות, השימוש ב-AI עשוי לא רק לשפר את איכות הביקורת אלא גם להניב חיסכון משמעותי בעלויות. התפתחות המבנה הארגוני הזה אינה היפותזה ספרותית. על פי דוח שוק העבודה של לינקדאין מינואר 2026 ("Building a Future of Work That Works"), בשנתיים האחרונות נוצרו ברחבי העולם כ-1.3 מיליון משרות חדשות הקשורות ל-AI, בהן תפקידי גשר חדשים בתחומים כמו Forward-Deployed Engineer, שצמח פי 42 מאז 2023, AI Engineer, שצמח פי 13. במקביל, בדוח "Work Change Report" של לינקדאין מינואר 2025 נמצא כי מספר החברות בארצות הברית המעסיקות עובד בתפקיד ראש תחום בינה מלאכותית גדל פי שלושה בחמש השנים האחרונות וב-50% נוספים בשנתיים האחרונות בלבד. התפקידים הללו הם המערכת החדשה של בקרה על מערכות אוטונומיות.

מתווה אופרטיבי לשילוב בינה מלאכותית בעבודת הביקורת

הפרק הקודם ארגן את הצעדים לפי לוח הזמנים שבו ייושמו. פרק זה מארגן אותם לפי הצירים התוכניים בעבודת הביקורת. פרק זה משלים את האמור עד כה במתווה אופרטיבי מפורט שנשען על מערכת שיקולים נורמטיביים ואופרטיביים. המתווה נבנה סביב תשעה צירי פעולה המכסים שני מעגלים משלימים: מעגל כנימי - הטמעת בינה מלאכותית בתוך מוסד הביקורת עצמו, ומעגל חיצוני - פיתוח יכולת הביקורת על בינה מלאכותית המשמשת את הגופים המבוקרים. נטען כי שני המעגלים ניזונים זה מזה: מוסד ביקורת שאינו משתמש בכלים מתקדמים אינו יכול לבקר ביעילות גוף שמשתמש בהם.

McKinsey Global Institute, *A New Future of Work* 69, לעיל ה"ש 33. הדוח מצביע על 12 מיליון שינויי תעסוקה נדרשים באירופה עד 2030.

70 eDiscovery הוא תהליך מובנה לאיתור, סינון, ניתוח ושימור מסמכים דיגיטליים רלוונטיים לחקירה או להליך משפטי. הוא מבוצע באמצעות פלטפורמות ייעודיות כגון Everlaw ו-Relativity וכיום לתקני EDRM (Electronic Discovery Reference Model).

Organisation for Economic Co-operation and Development, OECD AI Principles (2019 updated 2024), <https://www.oecd.org/en/topics/71-ai-principles.html>. ai-חלטת ה-OECD מגדירה חמישה עקרונות לאימוץ AI אחראי במגזר הציבורי.

ההכשרה המקצועית. המודל המוצע כולל צוותים היברידיים שבהם למבקר אנושי אחד יש עשרות סוכני AI הפועלים תחתיו במקביל⁷³.

הדרך הפחות טראומטית למעבר היא הימנעות מצמצום כוח האדם הקיים והסתמכות על תחלופה טבעית. בהערכה שמרנית, כ-3% - 5% מכוח האדם של מוסד ביקורת ממוצע עוזבים בכל שנה בשל פרישה, מעבר לסקטור הפרטי או קידום בין-ארגוני, בדומה לשיעורי תחלופה טיפוסיים בשירות הציבורי הישראלי ובדמוקרטיות מפותחות⁷⁴. על פני חמש עד עשר שנים, עצירת גיוס חלקית, למשל גיוס עובד אחד במקום כל שני עובדים שעוזבים, עשויה לצמצם את הצוות האנושי ב-15% - 30% בלי לפגוע בעובדים הקיימים. מהלך זה מחייב תכנון אסטרטגי ארוך טווח.

סיכום

מאמר זה פתח בשאלה אם העין הפקוחה חייבת להיות עין אנושית. התשובה שעולה מן הנתונים, מתרחישי השימוש בעולם ומן הניתוח הנורמטיבי ברורה מכפי שרבים היו רוצים להודות - לא, היא אינה חייבת להיות אנושית, וייתכן שהגיעה העת לומר בכנות שיש צורך בהתאמה ובהכנה משמעותית ליום זה.

הטכנולוגיה הקיימת כבר היום - לא בעוד עשור ולא בעוד חמש שנים - מסוגלת לבצע את רובו המכריע של תהליך הביקורת: איסוף הנתונים, זיהוי החרגות, הצלבת המקורות, גיבוש הממצאים וניסוח הדוחות. הפער בין הפוטנציאל, הנאמד ב-80 עד 94 אחוזים בתחומים הרלוונטיים, ובין השימוש בפועל, שאינו עולה על 15 עד 33 אחוזים, אינו פער של יכולת אלא של מוכנות מוסדית. יכולות הסוכנים הדיגיטליים מוכסלות מדי חודשים ספורים, כך שהפער הזה יצטמצם בין אם מוסדות הביקורת יערכו לכך ובין אם לאו. היעדר היערכות אינו עיכוב של השינוי, הוא רק שינוי הזהות של מי שיוביל אותו.

אולם, חשיבותו של המאמר איננה בתיאור המצב הטכנולוגי - היא בהיפוך שאלת המוצא. כל עוד הדיון הציבורי נסוב סביב השאלה "כיצד לשלב את הבינה המלאכותית בעבודת

פיילוט במקביל לעבודה האנושית הרגילה כדי להשוות ממצאים; תיעוד מדויק של שיעור השגיאות, שיעור הזיהויים כוזבים ושיעורי ההחמצה; והכפפת כל שלב לביקורת איכות מקצועית בהתאם לתקני ISSAI ולמדריך לביקורת המדינה.

תחומי ליבה שיישמרו כולם או בחלקם בידי אדם

חרף הפוטנציאל הרחב של הבינה המלאכותית, קיימים תחומים שבהם שיקול הדעת האנושי הוא תנאי לקיום תקין של הביקורת. מדובר בביטחון, בענייני חוץ ומודיעין ובכל תחום שהכרעה בו מחייבת איזון בין זכויות יסוד (כגון פרטיות, הליך הגון או חופש ביטוי) לבין אינטרסים ציבוריים. בתחומים אלה ייקבע העיקרון "human-in-command", קרי: המבקר האנושי מגבש את המסקנה הסופית, אך נעזר ב-AI לשם עיבוד חומרי הגלם. תקנת ה-EU AI Act מציעה מסגרת תקדימית שניתן להתאים לצורך זה.

רגולציה על בינה מלאכותית הפועלת ללא בקרה אנושית

כאשר תחומי פעולה שלמים של הביקורת יעברו לידי מערכות אוטונומיות, תידרש רגולציה ייחודית להפעלתן: חובת תיעוד שלבי ההסקה, חובת הסבריות, חובת עדכון הגוף המבוקר בקיומה של ביקורת אוטונומית וחובת ההותרה של זכות ערעור אנושית במקרה של ממצא שלילי. מסגרת ה-NIST AI Risk Management Framework ותקנת ה-EU AI Act מספקות בסיס לעקרונות אלה. חקיקה ייעודית בישראל, שתבוא בהשראת חוק הביקורת הפנימית וחוק-יסוד: מבקר המדינה, תידרש לתקן את הדין כדי להתאימו לעידן של ביקורת נטולת יד אדם⁷².

אסטרטגיית משאבי אנוש חדשה למוסדות ביקורת

המעבר לעידן של ביקורת נתמכת AI מחייב עדכון מקיף של אסטרטגיית משאבי האנוש במוסדות ביקורת. נדרשות מיומנויות חדשות - הנדסת פרומפטים, פיקוח על סוכני AI, ניתוח נתוני עתק ובדיקת מודלים. יש לבנות מסלולי הסבה מקצועית לעובדי ביקורת, להגדיר תקנים ייעודיים של "מפקח על סוכני AI" או "מבקר מומחה לבינה מלאכותית" ולשנות את תוכניות הלימוד באוניברסיטאות ובמוסדות

72 National Institute of Standards and Technology (NIST), *AI Risk Management Framework (AI RMF 1.0)*, NIST AI 100-1 (2023), <https://www.nist.gov/itl/ai-risk-management-framework>. מסגרת מתודולוגית לניהול סיכונים בינה מלאכותית המשמשת בסיס לרגולציית AI בארצות הברית ובגופי ממשל.

73 Claude, Anthropic, *Labor Market Impacts of AI* 73, לעיל ה"ש 6. המחקר ניתח מיליוני שיחות אנונימיות עם המודל Claude ומדד שימוש ב-AI בקטגוריות תעסוקה שונות של משרד העבודה האמריקני.

74 OECD, *Government at a Glance 2025* (2025), https://www.oecd.org/en/publications/government-at-a-glance-2025_0efdbcd-en.html. פרק 5 על תעסוקה במגזר הציבורי. הערה: שיעור התחלופה השנתי בשירות הציבורי המרכזי במדינות OECD נע בדרך כלל בטווח 3% - 7%, בהתאם להבדלים מבניים, מבחינת פנסיה והיצע שוק העבודה.



ואולי אף יתעצם דווקא משום שיופנו אליו כחות עבודות טכניות שמסיטות אותו מהמקום שבו הוא באמת יוצר ערך. הגדרת המבקר האנושי של העתיד דרך השכבות הללו בלבד איננה פשרה על המקצוע; היא מימוש המלא הראשון של הערך שהמקצוע התיימר לייצג מלכתחילה.

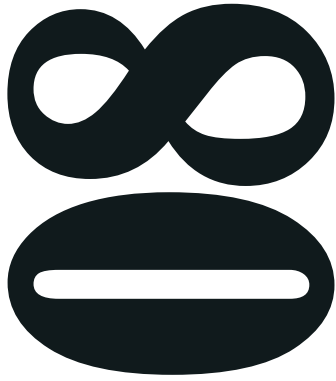
עבור ישראל - שמבקר המדינה שלה ניחן בסמכויות חוקתיות רחבות באופן יוצא דופן, ושהאקו-סיסטם הטכנולוגי שלה נמנה עם המתקדמים בעולם - השאלה אינה אם להיערך אלא כמה מהר לעשות זאת. המוסדות שיקדימו לא רק ישמרו על רלוונטיות - הם יגדירו מחדש את הסטנדרט שלפיו תימדד ביקורת המדינה בעשורים הבאים. המוסדות שימתינו יגלו שהציבור שהם אמורים לשרת כבר אינו מוכן להסתפק בדוח שמתפרסם חמש שנים אחרי המחדל, בשעה שכל אזרח מבין שהטכנולוגיה יכולה לזהות אותו מראש.

כותרת המאמר - המבקר האחרון? - הציגה שאלה. מי שממהר לענות עליה בביטחון - "לא, זה לא יקרה" או "התפקיד רק ישתנה" - אינו מגונן על מקצוע הביקורת, אלא על דמות מסוימת של מבקר. הביקורת תשרוד; היא אף עשויה לשגשג כפי שלא שגשה מעולם. אך תפקיד המבקר האנושי, במתכונת שאנו מכירים כיום, עשוי בהחלט להסתיים. השאלה שנותרת - המרתקת והמטרידה כאחת - איננה כיצד למנוע זאת, אלא האם ראוי לנסות למנוע.

המבקר, הוא מאשר בשתיקה את ההנחה שהעבודה הזו, בליבתה, חייבת להישאר אנושית. המאמר טוען שההנחה הזו טעונה הצדקה ואינה מובנת מאליה. המצב הקיים - שבו חלקים מן המגזר הציבורי אינם מבוקרים, הדגימה חלקית והדוחות מגיעים לעיתים שנים לאחר מעשה - אינו תוצר של בחירה מושכלת כי אם של מגבלת המשאבים האנושיים. כאשר המגבלה הזו מוסרת, ההיצמדות לסטטוס קוו חדלה להיות ברירת מחדל והופכת להחלטה שיש להצדיקה במפורש.

הציבור שהביקורת משרת אינו חפץ בעבודתו של המבקר; הוא חפץ בתוצריה - באמונה שהכסף הציבורי מנוהל כראוי, שההחלטות מתקבלות בשקיפות ושהמחדלים מזוהים ומתוקנים בזמן שבו עדיין אפשר לתקנם. אם ניתן לספק תוצרים אלה בכיסוי מלא במקום בדגימה, ברציפות במקום ברטרופסקטיבה ובתוך ימים במקום שנים - הרי שהזכות הציבורית לקבלם גוברת על הזכות המקצועית לבצעם באופן שבו התרגלנו שיבוצעו. ההיצמדות למבקר האנושי אינה נאמנות לערך הביקורת; במקרים רבים היא הפכה לחסם בפניו.

אין בכך קביעה שהגורם האנושי הופך מיותר לחלוטין. יש היבטים של שיקול דעת - הכרעות ערכיות, התמודדות עם לחצים פוליטיים ופרשנות של ממצאים בעלי השלכות ציבוריות כבדות - שבהן תפקידו של האדם יישאר מרכזי,



ישראל כמובילת סייבר עולמית - הדרך, ההובלה וחיוניות הביקורת להמשך ההצלחה

יגאל אונא

מר יגאל אונא שירת במשך 33 שנה בשירות הציבורי-ביטחוני במגוון תפקידים, בין היתר כיהן כראש אגף הסייבר-סיינט בשב"כ וכראש מערך הסייבר הלאומי (2018 - 2022). לאורך כל שנות שירותו עסק בתחום לוחמת המידע והסייבר ההתקפי וההגנתי והיה מופקד על הנחיית האמונים על התשתיות הלאומיות הקריטיות של ישראל להתגוננות מפני איומי הסייבר ועל תפיסות הביטחון הלאומיות בתחום. מייעץ בנושא לממשלות ולדיקטוריונים בארץ ובעולם ומקדם את תעשיית הסייבר הישראלית כמובילה בעולם.

המאמר בקצרה

- במאמר נבחנות התפתחות תחום הסייבר ולוחמת המידע והשלכותיה על הביטחון הלאומי, הכלכלה והממשל בישראל בעידן הדיגיטלי, בדגש על מאפייני האיום, מגמות ההתעצמות של תקיפות סייבר ומשמעותן האסטרטגית עבור מדינות וארגונים.
- עלייתה של מדינת ישראל למעמד של מעצמת סייבר עולמית נובעת משילוב בין תפיסה לאומית, הון אנושי איכותי, אקו-סיסטם מתקדם ושיתוף פעולה בין גופי ממשל, מערכת הביטחון והתעשייה.
- לצד היכולות הטכנולוגיות והמבצעיות, למערך הביקורת - הפנימית והחיצונית - תפקיד מרכזי בהבטחת ניהול סיכונים אפקטיבי, בשיפור מתמיד של מערכי ההגנה ובחיזוק החוסן הארגוני והלאומי.
- שילוב הדוק בין הגנת סייבר מקצועית לביקורת עצמאית ושיטתית הוא תנאי הכרחי לשימור יתרונה היחסי של ישראל בתחום הסייבר ולהבטחת עמידותה מול איומים מתפתחים במרחב הדיגיטלי.

מבוא | הסייבר ולוחמת המידע - קווים לדמותם | כיצד מדינת ישראל מתמודדת עם איום הסייבר | כיצד הפכה ישראל למעצמת סייבר? | מה בין סייבר לביקורת? | סיכום

מבוא

ממש בסוף שנות השמונים של המאה העשרים), המרשתת (האינטרנט), הטלפונים הסלולריים, הטלפונים "החכמים" (שהם למעשה "מיני מחשב" יותר מאשר טלפון), לווייני ה-GPS, הדרכון הביומטרי, מטבעות הקריפטו, הענן, מערכות בקרה ממוחשבות (סקאדה), מרשתת הדברים (IoT), ה-Wi-Fi והתשלום דרך המובייל (NFC). בימים אלה אנו מתרגלים לשימוש בבינה המלאכותית (AI), והמחשוב הקוואנטי כבר נוקש על דלתנו.

בכל הטכנולוגיות הללו, גם "העתיקות" יותר, אנו משתמשים בדרך זו או אחרת; קשה לאנושות להיפרד מהמצאותיה המוצלחות. כך, גם תקציב התא המשפחתי שלנו לתקשורת ולדיגיטל רק הולך וגדל הדרגתית לאורך השנים עקב צריכה מוגברת ומגוונת.

קצב השינוי וההתפתחות הטכנולוגית משתנה במהירות. הקצב שבו טכנולוגיות חדשות מפציעות רק הולך וגובר, ואיתו עולה גם הסיכון.

בעולם הסייבר, כל טכנולוגיה כזו מספקת משטח תקיפה - כינוי למרחב המעניק לתוקף סייבר הזדמנות לבצע את זממו. משטחי התקיפה האלה רק הולכים ומתרחבים וחושפים את כולנו באופן חסר תקדים לסיכונים אינדיבידואליים וארגוניים מיידיים ברמות שונות של חומרה.

מהי תקיפת סייבר?

תקיפת סייבר משמעותה השגת שליטה ללא רשות במערכת מידע, או להפך, מניעת שליטה במערכת המידע. כאשר תוקף סייבר משיג שליטה, ביכולתו לבצע במערכת כל מה שירצה, לרבות העתקת מידע (ריגול), פגיעה במידע תקיפות כופרה (Ransomware), פגיעה במערכות עצמן עד כדי השמדתן, גנבת כסף או גרימת נזק בעולם הפיזי (למשל השבתת חשמל, הפלת מטוס או התכת כור גרעיני).

מערכת מידע היא כל דבר שיש בו מעבד נתונים. כיום מדובר בפריטים רבים המקיפים אותנו, בהם השעון, הטלפון ואפילו משקפיים ופריטי לבוש. הקישוריות של כמעט כל חפץ וכלי למרשתת מאפשרת לתוקף סייבר שנמצא עשרות אלפי קילומטרים מהקורבן לבצע את התקיפה כאילו היה בחדר הסמוך.

הפורום הכלכלי העולמי (The World Economic Forum; WEF), ששיא פעילותו הוא בכנס הנערך בדאבוס מדי שנה בשנה, מפרסם בכל שנה דוח סיכונים גלובלי. סיכון הסייבר "מכבב" שם בעשרת המקומות הגבוהים ביותר (מתוך יותר מ-30), ובשנים מסוימות הוא אף משוּבץ במקום השני או השלישי.

מטרת התקיפות

לתוקפי סייבר מניעים שונים ומטרות שונות. רוב רובן של התקיפות נועדו **למטרות כלכליות**: סחיטת כסף באמצעות הצפנת המידע של הקורבן ומניעת יכולתו להפעיל את

במסגרת תפקידי הממלכתיים הוריתי לספק לתשתיות הלאומיות הקריטיות בישראל הגנה מאיומי סייבר, משימה שדרשה ביקורים בשטח ופגישות עם הגורמים השונים בחברות. באחד מביקורי בתשתית קריטית כזו, עלתה שאלה שמעסיקה רבות את כל העוסקים בתחום: למי כפוף האחראי על הגנת הסייבר בארגון? לא ארויב כאן על הדילמות והשאלות המקצועיות הרבות (המצדיקות מאמר בפני עצמו), אולם מקובלים שני מודלים - כפיפות למנהל מערכות המידע (CIO, מנמ"ר) וכפיפות לאחראי הביטחון (קב"ט) - אף שיש יותר ויותר מקרים של כפיפות ישירה למנכ"ל עצמו.

בעת התלהטות הדיון המקצועי, בקע מקצה השולחן קול שקרא: "מדוע לא תחת מבקר הפנים?" - דממה השתררה. אני עטתי על ההערה כמוצא שלל רב ופיתחתי את הדיון והרעיון, שכן יש קווי דמיון רבים מאוד, שלא לומר התאמה, בין המשימות של שני בעלי התפקידים - ביקורת שוטפת, ניהול סיכונים, ציות ודיווח. גם מאפייני התפקיד הנדרשים - עצמאות, ישרה, סמכויות וכשירות מקצועית - חופפים.

האם התהליכים הגלובליים עוד יביאו אותנו לשינוי כאמור - העברת הכפיפות למבקר הפנים? מה נדרש לשם כך? ובכלל, מה בין הגנת סייבר לביקורת?

הסייבר ולוחמת המידע - קווים לדמותם

כדי לרדת לשורש המונחים "סייבר" ו"תקיפות סייבר" יש להרחיק עד לשחר ההיסטוריה. לאורך כל הקיום האנושי, כל טכנולוגיה שהמציאו בני האנוש לשיפור רווחתם ואיכות חייהם נוצלה במהירות רבה גם לרעה, כדי לפגוע בבני אנוש אחרים. ככל שטכנולוגיה כזו מלווה בהטעיה ובהונאה גדולות יותר, כך היא יעילה יותר. פרשת הסוס של טרויה תופסת מקום מיוחד בפולקלור הסייבר, שכן הרעיון להחדיר בעורמה מרכיב עוין ומסוכן בתוך אריזה תמימה של פסל עץ מהודר ומושקע לעורף החשוף של אויב, רעיון שהביא להכרעה במאבק צבאי, משמש עד היום כאנלוגיה נפוצה בעולם הסייבר, ו"סוס טרויאני" הוא כינוי רווח לפוגען סייבר המורכב מקוד זדוני המוחדר במעטפת תמימה למערכות הקורבן וכך מאפשר לתוקפים לבצע את זממם.

טכנולוגיות המידע - שם כולל לעולם הדיגיטלי ההולך ומשתלט על כל היבט ותחום חיים מודרני - פיצו לעולמנו אחרי מלחמת העולם השנייה, וככל שהתפתחו בקצב הולך ומואץ, כך התפתח גם הצד האפל והזדוני שלהן. למעשה, זהו הסייבר, בהגדרתו העממית המופשטת ביותר - ניצול לרעה של כל אותן טכנולוגיות.

בכמעט ארבעת העשורים שבהם אני פעיל בתחום חוויתי כניסה של כמה טכנולוגיות מידע: המודם ההיסטורי ששימש להעברת מידע בין מחשבים, הפקסימיליה (מהפכה של



האבולוציה של הסייבר

עולם הסייבר החל להתפתח עם כניסת המחשב הראשון, והתפתחותו הואצה כשהתאפשרה תקשורת בין מחשבים. השימוש המקובל במונח "סייבר" עצמו החל בסוף העשור הראשון של שנות האלפיים. בראשית התפתחותו הייתה התכלית המרכזית של הסייבר ריגול, קרי חדירה למחשב בחשאי ואיסוף מידע בלי שהקורבן מרגיש בכך. לאחר מכן החלו תקיפות למטרת גרימת נזק. הידועה שבהן היא Stuxnet שפגעה בצנטריפוגות העשרת האורניום באיראן בשנים 2009 - 2010. לאחר מכן החלו תקיפות הכופר - פגיעה מותנית (ההופכת לקבועה רק אם לא משלמים כופר). השלב ההתפתחותי הבא הוא תקיפות למטרות השפעה, ואנו אכן חווים זאת בעוצמה מאז מערכת הבחירות בארצות הברית ב-2016. סוג עדכני של תקיפות הוא תקיפות שמטרתן גנבת מטבעות קריפטו שהיקפן הולך ועולה. ההערכה המקובלת היא ש-80% מכלכלת צפון קוריאה מתבססים על ערוץ הכנסה זה.

תקיפות מוכרות

יש לציון כי עד כה לא אירע "11 בספטמבר סייבר". כלומר, לא אירעו אסונות בהיקף נרחב כתוצאה מאירוע יחיד. אולם יש מספר אירועים אשר ראויים לציון, שכן התעשייה העולמית כולה מתייחסת אליהם כאל "קו פרשת מים" בהתפתחות ההתייחסות לסייבר ולצורך להתגונן מפני איומיו. להלן נתייחס לשני אירועים כאלה.

יגואר-לנד רובר (JLR) - ענקית תעשיית רכב בריטית שספגה באוגוסט 2025 תקיפה ששיתקה אותה למשך שישה שבועות. הנזקים שהסב האירוע נחשבו לקטסטרופליים במונחים עסקיים:

1. **השבתה ארוכה** - שישה שבועות ללא ייצור של אף רכב חדש, במקום 1,000 רכבים חדשים בשבוע. כלומר אובדן הכנסה של 50 מיליון פאונד בשבוע.
2. **השפעה לאומית** - ממשלת בריטניה הקצתה ערבויות של 1.9 מיליארד פאונד, הכרה בכך שתקיפת סייבר על שחקן תעשייה מרכזי משמעה איום על הביטחון הלאומי.
3. **השפעה כלכלית מיידית** - הפסד מוערך בכ-2 מיליארד פאונד שהם יותר מכל הרווח הנקי של החברה בשנת 2024.
4. **היעדר ביטוח** - האירוע התרחש כאשר החברה הייתה לקראת אישור של פוליסה, ומשכך הוטל הנזק כולו על כתפי החברה.
5. **פגיעה חמורה בכל שרשרת האספקה** - החברה פרנסה עוד כ-100,000 נושאי משרות מחוצה לה רק בבריטניה. נוצר אפקט דומינו שפגע לאורכה ולרוחבה של השרשרת והביא עסקים רבים לחדלות פירעון.

מערכותיו, אלא אם ישלם סכום במטבע קריפטו עבור מסירת סיסמת הפיענוח. תחום זה התפתח בקצב אדיר מאז פרוץ טכנולוגית הבלוקצ'יין ויצירת עולם מטבעות הקריפטו האנונימיים, דבר המאפשר לעבריינים לסחוט דמי כופר באופן בטוח ויעיל.

חלק מרכזי נוסף של התקיפות תכליתו **דליף מידע** (Data Leakage), כלומר החצנה בפומבי של מידע - בעיקר פרטי ורגיש - של לקוחות, של מטופלים ועוד כדי לגרום למבוכה. גם כאן נכנס פעמים רבות מרכיב כלכלי בדמות סחיטת כסף מהקורבן, שעשוי להעדיף לשלם כדי להימנע מפירסום שעלול לחשוף אותו לתביעות משפטיות מצד לקוחותיו.

יש תקיפות שמטרתן **גרימת נזק** גרידא, בעיקר מצד מדינות וארגוני פרוקסי של מדינות. הן מכוונות לרוב נגד תשתיות קריטיות או גופים חיוניים (מערכות ותשתיות של חשמל, מים וכד').

זן ספציפי נפוץ של תקיפות הוא **תקיפות מניעת שירות**. תקיפות אלה מתמשכות בדרך כלל שעות בודדות או ימים בודדים, ותכליתן "סתימת" התקשורת במרשתת של הקורבן שיכול להיות יעד ספציפי או אפילו מדינה שלמה.

זהות התוקפים

מקובל לחלק את התוקפים לקטגוריות: מדינות, ארגוני טרור וקבוצות פשיעת סייבר, ולכל אחת מהן שם ידוע או כינוי שהצמידו לה חברות סייבר גדולות. יש גם בודדים שתוקפים, אבל אלה נדירים וזנחים. בעבר התוקפים היו גם אקטיביסטים והתארגנויות "אידיאולוגיות", אך אלה כמעט אינם קיימים עוד, והאינטרס הכלכלי או הלאומי גבר על כולם. עד היום, ללא יוצא מן הכלל, נטלו קבוצות פשיעת סייבר את האחריות לתקיפות משמעותיות, גם אם היה ברור לחלוטין שמאחורי התקיפה עומדת מדינה. כך למשל מוכרות הקבוצות "חנדלה" המזוהה עם איראן ו"הדרור הטורף" המזוהה עם האופוזיציה האיראנית. עד כה לא מוכרת נטילת אחריות של מדינה על תקיפת סייבר, והדבר נובע ממגוון סיבות, לרבות אימת הדין הבין-לאומי. מקרה בולט הוא תקיפת הסייבר נגד הממשל המקוון באלבניה ביולי 2022, שיוחס לממשלת איראן והביא לניתוק היחסים הדיפלומטיים בין המדינות ואף לפנייה לנאט"ו (שאלבניה חברה בה) בדרישה להפעיל את סעיף 5 לאמנה ולרתום את כלל המדינות החברות כנגד התוקפן - איראן. הפנייה הזו נדחתה בטענה שאין לאלבניה הוכחה מוצקה לכך שאיראן היא העומדת מאחורי התקיפה. אכן, תקיפות סייבר הן עדיין הפשע המושלם. בפועל, ניתן גם לזהות שילוב בין אינטרסים ומטרות. איראן למשל מגייסת לשירותיה קבוצות פשיעת סייבר פרו-איראניות ומסמנת להן מטרות ישראליות. התכלית היא גרימת נזק ומבוכה, אבל כל סכום כסף שהעבריינים מצליחים לסחוט מהקורבן נשאר בידיהם. מעין מודל Revenue Share.

(תזרים, יתרה, מדיניות עסקית וכו'), יודע אם בבעלותו פוליסת ביטוח סייבר ומנוסה בהצבת דרישות שאינן גבוהות מידי עבור הקורבן כדי שיעדיף לשלם. האמינות והמוניטין של קבוצות פשיעת הסייבר מזכירות עולמות עסקיים אחרים. הקורבן יודע שאם ישלם יקבל את התמורה, כי אחרת "שמו הטוב" של התוקף ייפגע אנושות. גם בתהליך המשא והמתן בין התוקף לקורבן שוררת התנהלות עסקית לכל דבר, ועסקים רבים בכל העולם מתייחסים לנזק כתוצאה מתקיפה כאל חלק בלתי נמנע כמעט של עולם העסקים.

5. **תקיפות ריסוס גנריות** - קבוצות הפשיעה פועלות, בדומה לארגוני מודיעין, גם בשיטות של "ריסוס", כלומר חיפוש רחב של יעדים אשר "תלויים נמוך מהעץ" וקלים יותר להשגה, גם אם הכיס שלהם אינו עמוק, בעיקר באזורי ה(SMB; Server Message Block), שם גם תקציבי הגנת הסייבר ויכולות ההגנה נמוכים במיוחד.

6. **ניצול פרצות וחולשות מפורסמות** - ברוב רובן של תקיפות הסייבר התוקפים משתמשים בשני סוגי כלים (והרבה פעמים בשילוב ביניהם): הגורם האנושי (שעליו יורחב בהמשך) וניצול חולשות. חולשה היא טעות אנוש בכתיבת קוד תוכנה של יצרנים, לרבות של חומרה (חברות סיסקו ואנבידיה למשל). מרבית הטעויות האלה אינן מזיקות בהיבט הסייבר, אולם מיעוטן בהחלט כן, ואלה עלולות להותיר פרצה חמורה. בדרך כלל החולשות הללו מתגלות בידי היצרנים עצמם או על ידי חברות הגנת סייבר שמוכרות את הגילוי ליצרנים באופן אחראי. במקרה כזה היצרן מפתח "חיסון" - טלאי (Patch) - תוכנה קטנה שמתקנת את החולשה. אולם, כדי להפיץ את החיסון לכלל הלקוחות שכבר שילמו נדרש פרסום פומבי רחב. מרגע שהדבר מתפרסם מתחיל מרוץ בין התוקפים שמייצרים כלי נשק סייברי שמנצל את החולשה לביצוע זממם ובין המגינים שנדרשים להתקין את החיסון. המונח Zero day מתייחס לכך: מתחילים לספור ימים מרגע שהחולשה מתפרסמת ברבים, ועד אז זו חולשה שרק התוקפים נהנים ממנה, ועל כן יוקרתה ומחירה גבוהים מאוד. משך הזמן בין פרסום חולשה להשמשתה כנשק בידי התוקפים ירד מחודשים ליום אחד. במרוץ הזה הצד המרושע כמעט תמיד מנצח, אלא אם יש למגן פתרון הולם.

7. **קריפטו** - מטבעות הקריפטו וארנקי הקריפטו הפכו למשטח תקיפה ויעד מבוקש לתקיפות סייבר. מדובר ביקום דיגיטלי שלם עם הון עתק. קבוצות פשיעה רבות מכוונות את מאמציהן לתחום זה. הוא אנונימי, בלתי מדווח ונטול אסדרה, גן עדן גם למדינות כמו צפון קוריאה ולמדינות סוררות דומות.

כך נחשפה הפגיעות של מודל Just-in-time בייצור תעשייתי לכשל דיגיטלי מערכת.

Colonial pipeline - ענקית אנרגיה אמריקאית שמשנעת תוצרי דלק מזוקק ממפרץ אמריקה וטקסס, בעיקר לחוף המזרחי של ארצות הברית. החברה מעבירה את הדלק "עד המייל האחרון", כלומר משנעת את מכליות הדלק עד לתחנות הדלק. במאי 2021 שיתקה קבוצת תקיפה רוסית את המערכת המינהלתית של החברה (המערכת התפעולית שמשנעת דלקים יכלה להמשיך לעבוד ללא הפרעה, אבל לא היה ניתן לגבות תשלום). שיתוק החברה (שסירבה לשלם כופר של 2 - 4 מיליון דולר, הכנסתה הממוצעת בכמה שעות עבודה) הביא למחסור חמור בדלק, לרבות בדלק סילוני, לשיתוק כלכלי ותחבורתי ולהכרזת מצב חירום בידי הנשיא בידן.

חומרת איום הסייבר

כיום היקף הנזק שמסיבות תקיפות סייבר זכה להגדרה "הכלכלה השלישית בגודלה בעולם", אחרי ארצות הברית וסין (ולכני הודו). הנזק מוערך בסכום דו-ספרתי של טריליוני דולר ארצות הברית. קצב הגידול עצמו גבוה - כ-15% - 20% בשנה. היקפן העצום של התקיפות נובע מכמה סיבות:

1. **היעדר אכיפה יעילה** - מעטים מאוד האסירים שנשפטו ונכלאו, כמעט בכל מדינה, על עבירות סייבר בלבד. היכולת לשבת במדינת הבית ולתקוף במדינות אחרות, חלקן גם ללא הסכמי הסגרה, ייחודית לתחום הסייבר. גופי האכיפה מאוגדים תחת ארגון האינטרפול, שרמת הבנת הסייבר אצלו נמוכה מאוד. גופי הגנת הסייבר ההולכים וצומחים בעולם מאוגדים גם הם ומשתפים מידע, אך הם אינם עוסקים כלל באכיפה ובהעמדה לדין, כך שמדובר בשני עולמות מקבילים שאינם נפגשים. במרווח בין העולמות הללו חוגגות כנופיות הסייבר.

2. **אנונימיות** - המרשתת והטכנולוגיה מאפשרות לתוקפים רמת הסתוות ואנונימיות גבוהה למדי (יכולת זו אינה עומדת בפני מדינות מתקדמות, אך אלה מתעכפות יכולות מול איומים לאומיים).

3. **פער משפטי** - הטכנולוגיה תמיד הקדימה את המשפט והאסדרה (רגולציה), והדבר מורגש היטב גם בפשיעת סייבר. קשה להוכיח עבירות סייבר גם כשהעברין מזהה ונתפס, וכאמור הזהויה והתפיסה מורכבים דיים. גם כשכבר מצליחים לספק הוכחות, בתי המשפט אינם מבינים את הנושא לעומק והעונשים בדרך כלל קלים יחסית.

4. **התנהלות כלכלית ועסקית של התוקפים** - לתוקף שמטרתו כלכלית יש כלים המאפשרים לו ללמוד את הקורבן מבעוד מועד. הוא מכיר את יכולותיו הכלכליות



בטרם אפרט את האינדיקציות והסיבות לכך, חשוב להבין כיצד הגענו עד הלום.

מדינת ישראל הייתה החלוצה העולמית של התייחסות ממשלתית מסודרת לאתגר ההגנה מפני איומי סייבר, שנים לפני שהמונח נכנס לשימוש רווח. כבר ב-1997, במסגרת פרויקט תהיל"ה (תשתית הממשלה לעידן האינטרנט), שולב מרכיב ההגנה על חיבור משרדי הממשלה למרשתת כחלק אינטגרלי מהאחריות ומהפעילות. ב-2002 החליטה הממשלה בהחלטה ב/84, באמצעות ועדת השרים לביטחון לאומי ("הקבינט"), להסדיר את הגנת המערכות הממוחשבות החיוניות במדינה והטילה על השב"כ את האחריות לנושא. כך הוקמה רשות אבטחת המידע - רא"ם - יחידה אינטגרלית בשב"כ, באגף אבטחה דאז, שעליה הוטל הביצוע הלכה למעשה, וישראל טבעה לראשונה את המונח תשתיות מידע לאומיות קריטיות (תמ"ק), תשתיות שעליהן מוחלות הנחיות של קצין מוסמך מטעם הממשלה, הן במגזר הפרטי הן במגזר הציבורי.

8. **ניצול הגורם האנושי** - כאמור, הגורם האנושי הוא אחד משני הגורמים המרכזיים לתקיפות סייבר. קצרה היריעה, אולם המשתמשים האנושיים הם משטח התקיפה האולטימטיבי. כך ניתן לנצל את הטעויות שאנחנו עושים, את ההטעויות שעובדות על כולנו, בחירת סיסמאות גרועה (כי קשה לזכור, כי מתעצלים להחליף וכו') או מצגי שווא יצירתיים (פישנינג). כמעט שאין בן אנוש שחסין למניפולציות או סתם לביצוע טעויות שמשמעותן פתיחת דלת לתוקף והורדת ההגנות אם הן קיימות.

כיצד מדינת ישראל מתמודדת עם איום הסייבר

כיום מדינת ישראל היא מעצמת סייבר במגוון פרמטרים, שנייה אולי רק לארצות הברית, ובפרמטרים מסוימים, כשמחשבים פר-גולגולת, אף קודמת לה.

חברות ישראליות הפועלות בתחום את רף ה-66 מיליארד דולר, לאחר שבשנת 2024 הגיע "רק" ל-5.8 מיליארד. עסקאות ענק של 32 מיליארד דולר ו-25 מיליארד דולר הפכו לחזון נפוץ.

כיצד הפכה ישראל למעצמת סייבר?

ישראל הפכה למובילה עולמית בסייבר לא רק בשל החלוציות בהתייחסות הממשלתית והלאומית המסודרת להגנה, אלא גם הודות למאפיינים רבים, שאנסה לסכם את העיקריים שבהם בקצרה.

1. **חוק שירות ביטחון** - בישראל קיים גיוס חובה של בני ובנות 18 לצה"ל. עד שנות האלפיים העדיפות הראשונה בניתוב המתגייסים האזרחיים לצה"ל הייתה הפנייתם לקורס טיס וליחידות עילית לוחמות. עם עליית הסייבר חל שינוי דרמטי בכך, וזכות הבחירה הראשונה עברה ליחידות הסייבר השונות, בראשן 8200. כך, בני נוער שחלקם צברו כבר יותר מ-10 או אף 12 שנות ניסיון בפעילות בתחום המחשוב, לפעמים על סף החוקית (ולא אציין מאיזה צד של הסף), שולבו עם גיוסם ביחידות הסייבר, שאליהן נבנו מסלולי הכשרה רבים ומגוונים שהשכילו גם לפלח בכילוחי משנה את הכישורים ואת היכולות של המתגייסים הצעירים. כך השכיל צה"ל לאתר את הטובים שבטובים בסייבר באוכלוסייה, ואז לתת להם "רישיון" והכשרה לבצע פעילויות שנחשבות עד היום פורצות דרך וכמעט דמיוניות, כולל בהחתמה לשירות קבע. לאחר ארבע או חמש שנות התנסות מבצעית השתחררו מרבית הצעירים משירות, ובשונה מרוב חבריהם לשירות הצבאי, גילו שניתן לרתום את הניסיון הרב והידע שצברו לכדי הצלחה בעולם העסקי. מימוש החזון התנ"כי "וכיתתו חרבותם לאתים" מעולם לא קיבל משמעות כה מוחשית. ניתן להסב את יכולות תקיפת הסייבר ולפתח בעזרתן "חיסונים ותרופות" - יכולות הגנה שיגנו מפני האיומים בתחום - ועוד תמורת תשלום מכובד. כך קמה בתחילת שנות התשעים של המאה העשרים חברת צ'קפוינט, שהקימו קבוצת יוצאי 8200 שלא עסקו כלל בהגנה בעת שירותם הצבאי, אבל השכילו לבנות פתרונות הגנה מתקדמים אל מול האיומים שהם עצמם היו שותפים לבנייתם. בעקבותיהם באו עוד ועוד, בכל מחזור שהשתחרר מצה"ל. יודגש כי אליהם הצטרפו גם בוגרי שירות צבאי שאינו בתחום הסייבר ואף אינו טכנולוגי, מפקדים ולוחמים מכלל הזרועות שחלקם למדו טכנולוגיה באקדמיה. אלה הוסיפו ממדים וכישורים עסקיים חשובים לתמהיל האמור לעיל.

ראש הממשלה בנימין נתניהו נחשף בשנת 2010 להתפתחות תחום הסייבר ההתקפי ביחידה 8200 ובכלל מערכת הביטחון הישראלית. אחרי שגם קרא את ספרו הבדיוני של פיליפ ג'י גולדסטיין "בבל שעת אפס", המתאר תרחיש של מלחמת עולם בין ארצות הברית לסין המשלב סייבר עם נשק קונבנציונלי, הטיל ראש הממשלה על פרופסור יצחק בן ישראל להקים מיזם לאומי בתחום הסייבר - המיזם הקיברנטי הלאומי - שהקיף את כלל הארגונים והיבטי העשייה השונים בתחום, אזרחיים וביטחוניים, הגנתיים והתקפיים. עבודת המטה המקיפה ארכה כשנה, ובסיומה הוגשו המלצות לממשלה, והיא קיבלה אותן בהחלטה 3611 באוגוסט 2011, תחת הכותרת "קידום היכולת הלאומית במרחב הקיברנטי". התוצאה המיידית הייתה הקמת מטה הסייבר הלאומי בכפיפות לראש הממשלה. המטה החל מייד בפעולות שונות למימוש המלצות אלה.

אחת ההמלצות המרכזיות הייתה הקמת גוף חדש, מבצע, שיבצע בעצמו פעולות שונות להגנה על כלל המרחב הקיברנטי הלאומי והקמה של גופי הגנה מגזריים בקרב כלל משרדי הממשלה וגופי ממשל נוספים כבנק ישראל (החלטות הממשלה 2443 ו-2444 בפברואר 2015). הרשות הלאומית להגנה בסייבר קמה בפועל באפריל 2016, וזמן קצר לאחר מכן קיבלה לידיה את רוב תחום האחריות מרא"ם בשב"כ. אולם ריבוי הגופים וחפיפת תחומי האחריות בין המטה הלאומי, רשות ההגנה, שב"כ והממונה על הביטחון במערכת הביטחון (מלמ"ב) לא אפשר התנהלות תקינה. באורח פלא הממשלה הבינה זאת בתוך פרק זמן קצר יחסית, כשנה וחצי, והמטה והרשות מוזגו לגוף אחד - מערך הסייבר הלאומי. גוף זה פועל החל בינואר 2018 ועד היום, בכפיפות ישירה לראש הממשלה, אך עדיין ללא חוק שמסדיר את פעילותו. הצעת החוק הראשונה פורסמה ביוני 2018, ולאחרונה, ביוני 2026, פורסמה הצעת חוק מתוקנת, כך שרוב הבסיס החוקי לפעילות המערך נובע מתיקונים ותוספות לחוק הסדרת הביטחון ולחוקים אחרים.

אולם בד בבד עם התהליכים הממשלתיים והבירוקרטיים החשובים הללו, קרה תהליך משמעותי לא פחות - עלייתה של תעשיית הסייבר הישראלית. את חוזקה של התעשייה הזו בקנה מידה עולמי ניתן להמחיש בעזרת היקף ההשקעות הגלובליות בישראל והיקף ההנפקים (אקזיטים) של חברות בישראל.

יותר מ-40% מההשקעות הגלובליות בתחום הסייבר מגיעים לחברות ישראליות. נתון זה מרשים עוד יותר בהשוואה לשנת 2019, אז "רק" 20% מההשקעות הגיעו לישראל, שכזכור, כוללת רק 0.1% מאוכלוסיית העולם.

הנפק (אקזיט) הוא מימוש ההשקעות בחברה (בדרך כלל בתחום ההיי-טק), כלומר הנפקה או רכישה בידי שחקן גדול יותר. בשנת 2025 לבדה חצה היקף ההנפקים של

2. **תרבות ההיי-טק הישראלית** - היעדר משאבי טבע, תחושת מחסור ומצור, מערכת חינוך טובה יחסית (גם אם מושמצת) וחשוב מכול - תרבות המעודדת יזמות, חדשנות, יצירתיות ואלתור וכן סלחנות כלפי כישלונות לצד נשיאה על כפיים של המצליחים וההישגיים - כל אלה דחפו ליצירת "אומת הסטארט-אפ". סף הכניסה להקמה של חברת היי-טק בכלל וחברת סייבר בפרט נמוך עד כדי הסתפקות בידע ובכמה מחשבים. ניתן אפילו להמיר משרדים בבתי קפה.

3. **הלכידות הישראלית** - בסייבר לא ניתן להצליח לבד. לא סוכנות בודדת ואף לא מדינה בודדת. זהו משחק קבוצתי מאין כמוהו בשל טבעו של מגרש המשחקים - המרשתת וטכנולוגיות המידע חובקות העולם. בודדים או גופים מנותקים לעולם לא יצליחו להתמיד בהתמודדות מוצלחת עם הקצב והמגוון של האיומים בתחום. שיתוף ידע, שיתוף מידע קונקרטי ואף שילוב יכולות שונות ומשלימות (שילוב האצבעות השונות לאגרוף) - כל אלה חיוניים להצלחת ההגנה מסייבר. לחברה הישראלית, על אף הביקורת העצמית המפותחת שלנו (וביקורת היא דבר חשוב, זו בדיקת הבמה לציין...) היא בעלת לכידות והדדיות מהגבוהות בעולם, בוודאי כשהדבר נוגע להתמודדות מול איומים. בשנת 2021 הוענק פרס ביטחון ישראל על ידי נשיא המדינה "ל"אורגן" - מנגנון המשלב שישה גופי ביטחון במאמץ מרוכז להגנת הסייבר על המדינה: מערך הסייבר הלאומי, צה"ל (יחידה 8200 באמ"ן וחטיבת ההגנה בסייבר באגף תקשוב), שב"כ, המוסד ומלמ"ב במשרד הביטחון. האורגן פעיל מבצעית כבר משנת 2013, והפרס הוענק לו לא רק על היכולות הטכנולוגיות שפיתחו הגופים במשותף, אלא בעיקר כהערכה על היכולת לפעול באופן משולב על אף ניגודי עניינים וקונפליקטים מובנים בין הגופים. לדוגמה, גופי המודיעין חרדים לשמירת ביטחון מקורות המידע ואינם מעוניינים לחשוף מודיעין המתקבל ממקורות רגישים, ואילו גופי הגנת סייבר משתמשים באופן גלוי ולעיתים "רועש" במידע כדי לייצר את ההגנה הנדרשת, ובכך מסכנים במודע את ביטחון המקורות. נוסף על כך, גופי סיכול פועלים להגעה ל"ראש הנחש" ואוספים מידע בסבלנות עד שהתמונה השלמה והמדויקת מתקבלת וניתן לפעול על פיה, ואילו גופי הגנה מנהלים את הסיכון באופן הדוק יותר ואינם דוגלים במתן הזדמנות לתוקף לבצע את זממו, אלא בסיכול פעילותו ברגע שבו היא מתגלה. מדוע האורגן כה מצליח על אף נתוני פתיחה מורכבים אלה? התשובה היא בתרבות המקומית של ה"ביחד" וההבנה שאף אחד מששת הגופים, גדול, חזק, וסמכותי ככל שיהיה, אינו מסוגל לנצח לבדו. אצוין כי עד היום מערכות ביטחוניות ברחבי העולם עומדות משתאות אל מול הצלחת

האורגן הישראלי, ללא כל שכפול או חיקוי מוצלח של המודל.

4. **האקו-סיסטם** - מדינת ישראל השכילה לייצר מערכת יעילה למדי של אקו-סיסטם בסייבר, המורכבת מפעילות הממשלה, האקדמיה והתעשייה האזרחית. מדוע אקו-סיסטם? כי בדומה למערכות דומות בטבע, המערכת מזינה את עצמה ודואגת להמשכיות עצמית ואף לצמיחה. הממשלה פתחה שבעה מרכזי מצוינות בסייבר במימון ממשלתי נכבד, בכל אוניברסיטה שהביעה עניין בכך. האקדמיה דואגת להכשרה ולליטוש של כוח אדם איכותי שנמשך לתחום בשל הפוטנציאל העסקי הגבוה והעניין המקצועי, התעשייה נהנית מכוח האדם הזה שמעשיר אותה ומעצים את יכולתה לייצר עסקים, ההון חוזר לממשלה במיסים, וגם האקדמיה נהנית מהתשומות במישרין ובעקיפין בדמות יוקרה ורלוונטיות. במילים אחרות, זהו מעגל קסמים חיובי ומתעצם שבו הממשלה מעודדת באמצעות גופים כמו מערך הסייבר הלאומי, משרד הכלכלה ומשרד הביטחון את התעשייה והאקדמיה לבנות ולשפר יכולות ואף למכור אותן לגופי ממשל ולקוחות בחו"ל.

5. **"עם כלביא יקום"** - הנסיבות האזוריות וההיסטוריה המדממת יצרו במדינת ישראל חברה אסרטיבית ולוחמנית באופייה, הבנויה למאבק מתמשך, שנלחמה לאורך כל שנות המדינה ואף לפני הקמתה על קיומה ועצמאותה. הנוער הישראלי גדל וצומח לאווירה ביטחונית והתמודדות מתמדת מול איומים (אפילו בעת נסיעות לחופשה בחו"ל, הצנעת סממנים ישראלים היא "צמצום משטח תקיפה" מסוג מסוים). הצעירים חווים זאת גם באופן אקטיבי, "ציד" של האיומים בעיקר במסגרת השירות הצבאי. לוחמת סייבר, הגם שעיקרה בעולם האזרחי-פילי, היא עדיין לוחמה שדורשת "ציד של הרעים", תחכום, תחבולה ובעיקר היעדר נאיביות ומצב נפשי של תחרות שאינה ספורטיבית. ברוב המדינות האחרות, שלשמחתן אינן חוות את האיום ברמה היומית, התמימות המבורכת היא גורם מעכב המונע התפתחות כישורים החיוניים לתחום הסייבר. בישראל זהו יתרון תחרותי במלוא מובן המונח.

מה בין סייבר לביקורת?

כפי שצינתי באנקדוטה בפתיחה, יש קווי דמיון רבים מאוד בין ביקורת להגנת סייבר.

תפקיד הביקורת בארגון הוא לספק הערכה של תהליכים פנימיים, להצביע על פערים וכשלים, לוודא ניהול סיכונים וציות לרגולציות, לזהות פערים ובעיות ממשליות ולדחוף לשיפור תהליכים ולהתייעלות. כל אלה מופיעים גם בכתב האחריות של מגן הסייבר בארגון.

להיותנו מהמדינות המותקפות ביותר בסייבר בעולם, אם לא המותקפת ביותר פר-גולגולת. להבנתי, על בסיס ניסיון רב-שנים, זהו חלק בלתי נפרד מהצלחתה של ישראל בתחום הסייבר ומהתבססותה כמעצמה עולמית בתחום. בשנים האחרונות זהו נושא ביקורת מרכזי, ומאז פרוץ הסייבר לחיינו פורסמו עשרות דוחות בנושא, מקצתם עוסקים רק בהגנת סייבר ואבטחת מידע ומקצתם מוקדש פרק ייחודי לנושא הסייבר. במשרד המבקר בוצעו שינויים ארגוניים כדי לשפר את היכולת המקצועית והקיבולת של המשרד בתחום, ואני גאה לציין שבתפקידי כראש מערך הסייבר הלאומי יכולתי לסייע מקצועית בתהליך המבורך. המבקר בדק, והוא חוזר ובודק את המוכנות ואת רמת הגנת הסייבר של הרשויות המקומיות (מספר פעמים), מערכת הבריאות, התשתיות הקריטיות, המידע הממשלתי, מערכות חירום לאומיות ועוד. בדיקות וביקורות אלה מציבות תמונה שקשה לחלוק עליה ודוחפות את הארגונים לבחינה עצמית ולתיקון שוטף של ליקויים ופרכות. הביקורת תורמת גם להעלאת המודעות והקשב של ההנהלה הבכירה לתחום ובכך גם לשיפור המקצועיות ולהגעה לתוצאות טובות יותר באופן שיטתי. בשונה אולי מחלק מהביקורות, חלק ניכר מהארגונים מבינים את החשיבות שבקבלת ממצאי הביקורת הנוגעים לתחום הסייבר, גם אם אינם מחמיאים, ופועלים במהירות לתיקון ולשיפור כל מה שנדרש.

סיכום

הסייבר הפך לאחד האיומים המרכזיים על מדינות, ארגונים וחברות בעידן הדיגיטלי. התלות בטכנולוגיה והקישוריות הגלובלית מרחיבות את משטחי התקיפה ואת פוטנציאל הנזק הכלכלי, התפעולי והלאומי, וזאת בקצב גובר הנגרם מקצב התפתחות הטכנולוגיה הדיגיטלית ותלותנו בה. מדינת ישראל הצליחה לבסס מעמד של מעצמת סייבר עולמית בזכות שילוב ייחודי של תפיסה לאומית, הון אנושי איכותי, אקו-סיסטם מתקדם ושיתוף פעולה בין גופי ממשל, ביטחון ותעשייה. עם זאת, הצלחה זו מחייבת תחזוקה מתמדת של יכולות ההגנה והסתגלות לשינויים מהירים בזירת האיומים. הביקורת - הפנימית והחיצונית - היא מרכיב חיוני במערך ההגנה, שכן היא מאפשרת זיהוי פערים, חיזוק של ניהול הסיכונים ושיפור מתמיד של רמת המוכנות והחוסן הארגוני והלאומי. שילוב אפקטיבי בין הגנת סייבר מקצועית לבין ביקורת עצמאית ואיכותית הוא תנאי הכרחי לשימור יתרונה של ישראל בתחום הסייבר ולהבטחת ביטחונה, יציבותה הכלכלית ואמון הציבור במערכת.

הגנת הסייבר של ארגון מתחילה מסקר סיכונים כולל וממפיו כלל הסיכונים לארגון שיכולים להיגרם כתוצאה מתקיפת סייבר: פגיעה ברציפות התפקודית, פגיעה ונזק של ממש, נזק כלכלי, נזק למוניטין, זליגת מידע רגיש של לקוחות, חטיפת המידע למטרות כופר ועוד ועוד. על בסיס הסקר נבנית תוכנית פעולה שוטפת המתבססת על פי רוב על סוג של אסדרה מקובלת - מסגרות כמו NIS2, NIST, ISO 27001 ועוד, הכוללות סדרה ארוכה ומפורטת של בקורות (Controls) ועמידה בדרישותיהן מגבירה מאוד את הסיכוי שהארגון ינהל את הסיכונים באופן טוב ולא ייפגע מתרחישי סייבר.

הבקרה על עמידות הארגון בפני סייבר מתבצעת הן באופן שוטף ורציף באמצעות טכנולוגיות שונות, סנסורים הממוקמים בכל רחבי הארגון מדווחים על ניסיונות תקיפה ולפעמים גם מונעים אותם, הן במסגרת מבחני חדירה (Penetration tests) המבוצעים מפעם לפעם על ידי גורמי מקצוע מחוץ לארגון, כסוג של ביקורת חיצונית, ומדמים פעילות של תוקף סייבר אותנטי, אך ללא כוונות זדון וללא גרימת נזק בפועל.

מבדקי החדירה בתצורותיהם השונות הם המרכיב האיכותי ואולי החשוב ביותר במערך ההגנה של ארגון. עמידה בנהלים ובבקורות, הגדרה נכונה של טכנולוגיות ופתרונות ההגנה, אימון והכשרת העובדים - כולם מכוונים למטרה אחת: למנוע בפועל תקיפה מוצלחת. מבדק החדירה, אם הוא נעשה בצורה מקצועית וטובה, הוא "היהלום שבכתר" שבודק במבחן אמת את כלל המרכיבים השונים, בלי "עיגולי פינות" או הנחות. דומה הדבר לבניית צוללת והכנסתה למים כדי לוודא שאין בה שום נזילה.

וכמו בכל ביקורת, רגע אחרי שהסתיימה בדיקת החדירות האחרונה, הגיע זמן הבדיקה החוזרת לשם בחינת תיקון הליקויים שהתגלו (ותמיד מתגלים פערים בהגנת הסייבר שכן שילוב של טכנולוגיות ובני אדם חשוף במיוחד לפערים וכשלים) וחוזר חלילה.

הבדל אחד נעוץ אולי בפוטנציאל הכלכלי. יש כמה חברות סייבר ישראליות שמתמחות בתצורות וגרסאות שונות של מבדקי החדירה הללו. חברות אלה הגיעו לשווי של יותר ממיליארד דולר.

לכן, אין זה מפתיע שישראל היא כנראה המובילה בעולם גם בתחום הביקורת משולבת הסייבר. פירמות מובילות בעולם לראיית חשבון, ייעוץ וביקורת כמו דלווייט (Deloitte) ו-BDO הקימו בישראל את מרכזי הסייבר העולמיים שלהן, המיועדים למתן שירותי סייבר שונים לכלל לקוחותיהן ברחבי העולם.

משרד מבקר המדינה בישראל גם הוא מהחלוצים והמובילים בתחום בעולם, כראוי למעצמת סייבר וגם



ביקורת המדינה על שימוש בבינה מלאכותית בגופים ציבוריים: עקרונות לבחינה מוסדית ונורמטיבית

יובל ריינפלד וענבל בן ארצי

ד"ר יובל ריינפלד הוא מרצה באוניברסיטת רייכמן, יו"ר (משותף) ועדת בינה מלאכותית בלשכת עורכי הדין וחבר בפורום המומחים הלאומי לבינה מלאכותית.

עו"ד ענבל בן ארצי היא מנכ"לית HighLaw ועוסקת בפיתוח, בהדרכה ובהטמעה של כלי AI.

המאמר בקצרה

- שילוב בינה מלאכותית לצורך קבלת החלטות על ידי גופים ציבוריים טומן בחובו סיכונים לפגיעה בזכויות אזרח.
- בישראל לא קיים חוק רוחבי בעניין השימוש בבינה המלאכותית, והרגולציה היא מבוזרת וסקטוריאלית.
- הדין האירופאי והאמריקאי עשויים להוות השראה לאימוץ עקרונות לפיקוח על השימוש בבינה מלאכותית.
- בהיעדר מסגרת חקיקתית אופקית, ביקורת המדינה עשויה לתרום לא רק לחשיפת ליקויים, אלא גם לעיצוב סטנדרט ציבורי של שימוש אחראי בבינה מלאכותית.
- לשם כך נדרשת הקמת תשתית ביקורתית חדשה, הכוללת שפה מקצועית עדכנית, מומחיות טכנולוגית, שיתוף פעולה בין-מוסדי, כלי הנחיה מוקדמים ויכולת ללמוד מן הנעשה בעולם בנושא.

מבוא | מהות הסיכון: מערכות בינה מלאכותית ציבוריות אינן כלי ניטרלי | מסגרת נורמטיבית:
הדין הקיים ופערי הרגולציה | שישה עקרונות לבחינה מוסדית | תפקיד מבקר המדינה: לקראת
ביקורת מותאמת AI | סיכום

לנתונים ולתהליכים פנימיים שאינם נגישים בדרך כלל לגורמים חיצוניים; ושלישית, ניסיונו המצטבר בבחינת חוקיות, תקינות ויעילות של גופים ציבוריים מספק בסיס מוסדי מתאים להרחבת הביקורת גם לתחום השימוש בבינה מלאכותית.

גם במבט השוואתי ניכרת מגמה ברורה של התעוררות מוסדות ביקורת עליונה לסוגיה זו. משרד הביקורת הלאומי של בריטניה (UK National Audit Office; NAO) עסק בשימוש בבינה מלאכותית בממשלה הבריטית, משרד הביקורת הממשלתי של ארצות הברית (US Government Accountability Office) פרסם מסגרת עבודה לבחינת השימוש הממשלתי ב-AI, ובמדינות אירופיות שונות התחדד הקשר בין שימוש במערכות בינה מלאכותית לבין שאלות של שקיפות, זכויות ופיקוח. ישראל יכולה ללמוד מניסיונם של גופים אלה ולהתאים את כלי הביקורת לצרכיה.

חלק א' של המאמר עוסק במהות הסיכון הייחודי שמציבות מערכות בינה מלאכותית בשירות הציבורי; חלק ב' סוקר את המסגרת הנורמטיבית הקיימת ומצביע על פערי ההסדרה בישראל, בין היתר בהשוואה לדין האירופי; חלק ג' מציג שישה עקרונות לבחינה מוסדית של שימוש בבינה מלאכותית; וחלק ד' מתרגם עקרונות אלה לכלים ביקורתיים מעשיים ומציע המלצות מדיניות. המאמר מסתיים בסיכום שמחזיר את הדיון מן המישור הנורמטיבי אל משמעותו המעשית עבור הפרט אשר עשוי להיפגע מהחלטה מינהלית הנתמכת במערכת בינה מלאכותית.

מהות הסיכון: מערכות בינה מלאכותית ציבוריות אינן כלי ניטרלי

בשירות הציבורי פועלות מגוון מערכות בינה מלאכותית: מערכות ניבוי המזהות "פרופיל סיכון", מנועי דירוג לצורכי תיעודף טיפול, מסנני בקשות אוטומטיים ועוזרי החלטה לפקידים ולבעלי סמכות. חרף השונות ביניהן, למערכות הללו מאפיין יסודי משותף: אין מדובר בכלים מינהליים ניטרליים במובן הרגיל. אופן פעולתן, מקורות הנתונים שעליהם הן נשענות והמשקל שניתן לפלט שהן מייצרות

בשנים האחרונות החלו גופים ציבוריים ברחבי העולם לבחון שימוש במערכות בינה מלאכותית (להלן גם - AI) לצורכי דירוג סיכון, תיעודף טיפול ותמיכה בקבלת החלטות. במערכות רווחה, למשל, משתמשים לעיתים בכלים המעבדים מגוון רחב של נתונים, כגון היסטוריית דיווחים, מצב כלכלי, מבנה משפחתי ושינויים במקום המגורים, כדי להפיק "ציון סיכון" המשמש בסיס לקביעת סדרי עדיפויות ולעיתים אף להחלטות רגישות הנוגעות להתערבות בחיי משפחה. ניתן למצוא דוגמאות דומות לשימוש באלגוריתמים לצורך זיהוי צרכים עתידיים ותכנון תקציבי במגזר הציבורי, כגון הקצאת משאבים במערכת הבריאות, לרבות תכנון מיטות, כוח אדם וציוד¹. מציאות זו מעלה חשש כי כאשר אנשי מקצוע יפעלו תחת עומס כבד תגבר ההסתמכות על פלט המערכת, ואילו היכולת להסביר את ההחלטה, לנמקה באופן עצמאי ולאפשר ביקורת אפקטיבית עליה תיחלש².

אין מדובר בחשש תיאורטי בלבד. מערכות דומות כבר פועלות ברשויות רווחה, במחלקות הטבות סוציאליות ובמוקדי אכיפה במדינות שונות, והדיון בשילובן במגזר הציבורי הולך ומתרחב גם בישראל³. על רקע זה מתחדדת שאלה יסודית: כאשר גוף ציבורי משלב בינה מלאכותית בקבלת החלטות הנוגעות לזכויות האזרח, מהם הכלים הנורמטיביים והמוסדיים שעל ביקורת המדינה להפעיל כדי להבטיח שלטון חוק, אחריותיות ושמירה על זכויות? שאלה זו מקבלת משנה תוקף נוכח פערי ההסדרה הקיימים.

באיחוד האירופי נכנסה לתוקף מסגרת רגולטורית ייעודית, ה-EU AI Act, הקובעת דרישות מחמירות ביחס למערכות בינה מלאכותית, ובפרט למערכות המוגדרות כבעלות סיכון גבוה⁴. בישראל, לעומת זאת, טרם עוצבה מסגרת חקיקתית ייעודית ומחייבת בתחום זה, ונותרו פערי פיקוח ניכרים. ביקורת המדינה עשויה למלא תפקיד מרכזי בחלל זה.

מאמר זה טוען כי ביקורת המדינה היא אחד הגורמים המוסדיים המתאימים ביותר למילוי התפקיד. שלושה טעמים עיקריים תומכים בכך: ראשית, עצמאותו המוסדית של מבקר המדינה מאפשרת לו לבחון את פעולת הרשות המבצעת; שנית, סמכויותיו מקנות לו גישה למסמכים,

1 European Commission, *Artificial Intelligence in Healthcare* (n.d.), https://health.ec.europa.eu/ehealth-digital-health-and-care/artificial-intelligence-healthcare_en

2 UK Government, *Artificial Intelligence Playbook for the UK Government* (2025), <https://www.gov.uk/government/publications/ai-playbook-for-the-uk-government/artificial-intelligence-playbook-for-the-uk-government-html>

3 Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, And Punish The Poor* (2018).

4 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence, OJ L, 12.7.2024.



לא תמיד יכול להסביר לאזרח כיצד התקבלה התוצאה, ולעיתים גם מפתחי המערכת עצמם אינם מסוגלים להציג הסבר מלא וברור להיגיון הפנימי של המודל. אי-שקיפות זו, הנדונה תדיר בספרות כבעיית הקופסה השחורה (Black box), מאיימת במישרין על עקרונות יסוד של המשפט הציבורי, ובהם זכות הטיעון, חובת ההנמקה ועקרון שלטון החוק⁵.

כאשר אזרח מבקש לדעת מדוע בקשתו נדחתה, מדוע סווג באופן מסוים או מדוע עניינו תועדף באופן מסוים, ותשובת הרשות נשענת הלכה למעשה על פלט אלגוריתמי בלתי

עשויים להשפיע באופן ממשי על זכויות, על נגישות לשירותים ועל אופן הפעלת שיקול הדעת המינהלי. שלושה מאפיינים מרכזיים מבדילים אפוא את המערכות האלה מכלים מינהליים רגילים ומצדיקים פיקוח ייחודי עליהן, כפי שיפורט להלן.

אי-שקיפות אלגוריתמית

מודלים מבוססי למידת מכונה, ובפרט מודלים מורכבים, עשויים להפיק תוצרים שקשה לפרשם באופן אינטואיטיבי. פקיד המקבל ציון סיכון, דירוג או המלצה אופרטיבית

Frank Pasquale, The Black Box Society: The Secret Algorithms That Control Money and Information 1 - 18 (2015). 5

מוסבר, נפגעת יכולתו להבין את ההחלטה, לתקוף אותה ולהעמידה לביקורת.⁶

הספרות האקדמית מצביעה על כמה מקורות לאי-שקיפות בפעולתן של מערכות אלגוריתמיות, ובהם מורכבות טכנית המקשה על הבנת אופן הפעולה של המודל, סודיות מסחרית או קניינית המגבילה את היכולת לבחון את המערכת וכן היעדר גילוי מספק מצד הרשות בעניין האופן שבו נעשה בה שימוש.⁷ גורמים שונים אלה לאי-שקיפות עשויים להתקיים גם בהקשרים ציבוריים בישראל, ומשום כך עליהם לעמוד במוקד בחינתה של ביקורת המדינה.

הטיות מובנות

מערכות בינה מלאכותית מבוססות, במישרין או בעקיפין, על נתוני עבר. כאשר נתונים אלה משקפים דפוסים היסטוריים של אפליה, הקצאה בלתי שוויונית של משאבים או סטיגמטיזציה של קבוצות אוכלוסייה מסוימות, קיים חשש ממשי כי המודל לא רק ישחזר דפוסים אלה, אלא אף יעצים אותם.

הקושי המרכזי טמון בכך שההטיה אינה נובעת בהכרח מאפליה מכוונת במובנה הישיר, אלא מן העובדה שהמערכת מזהה קורלציות סטטיסטיות בתוך מציאות חברתית ומוסדית שכבר הייתה מוטה מלכתחילה ופועלת על בסיסן.

כך, למשל, אם בשל נגישות נמוכה, דפוסי טיפול בלתי שוויוניים או הטיות מוסדיות אחרות סווגה בעבר קבוצה מסוימת כבעלת "סיכון" גבוה יותר בהקשרים של זכאות לשירותי רווחה, של אכיפה או של גישה לשירותים ציבוריים, המודל עשוי לפרש נתונים אלה כאינדיקציה אובייקטיבית לכאורה לסיכון גם בעתיד. במובן זה, הטיה שיטתית אינה תוצר של כוונה זדונית בהכרח, אלא של מבנה הנתונים, של תנאי יצירתם ושל האופן שבו המציאות החברתית מתורגמת למשתנים חישוביים.

פוטנציאל לכשל אלגוריתמי בעל השפעה רחבת

כאשר גוף ציבורי מפעיל מערכת בינה מלאכותית בקנה מידה רחב, הסיכון הטמון בכך אינו מתמצה בהחלטה שגויה אחת, אלא ביכולתה של אותה שגיאה להשתכפל במהירות ובהיקף נרחב. בניגוד לטעות אנוש "רגילה", אשר לרוב מוגבלת למקרה מסוים, לבעל תפקיד מסוים או להקשר עובדתי נקודתי, טעות אלגוריתמית שיטתית עלולה להשפיע בעת ובעונה אחת על ציבור רחב מאוד של פרטים. יתרונה התפעולי של המערכת - יכולתה לפעול באופן

אחיד, רציף ומהיר - הוא גם מקור הסיכון הייחודי שלה: ככל שהמערכת יעילה יותר בהפקת החלטות או המלצות בקנה מידה רחב, כך גדל גם פוטנציאל ההפצה של כשל מובנה, הטיה שיטתית או הנחת יסוד שגויה.

הקושי מתחדד עוד יותר משום שפגיעה אלגוריתמית רוחבית לא תמיד נחשפת מייד. כל אזרח או תושב חווה את ההחלטה כעניינו הפרטי בלבד, ואינו יכול לדעת אם מדובר בשגיאה חד-פעמית או בתוצאה של דפוס שיטתי המשפיע באופן דומה גם על אחרים.

במילים אחרות, אף שהפגיעה נחוות באופן אינדיבידואלי, הכשל הוא לעיתים קרובות מבני וקולקטיבי. דווקא בשל כך, מנגנוני הביקורת הרגילים, הנשענים על תלונות פרטניות או על בחינה של מקרים בודדים, עלולים להתקשות בזיהוי הבעיה בזמן אמת. כאן מתחדדת חשיבותו של מוסד ביקורת בעל פרספקטיבה מערכתית.

גוף כגון מבקר המדינה אינו נדרש לבחון רק אם החלטה מסוימת הייתה סבירה בנסיבותיו של אדם יחיד, אלא אם קיים דפוס רחבי של קבלת החלטות לקויה, בלתי שוויונית או בלתי מוסברת. יכולתו לעיין בנתונים מצטברים, להשוות בין מקרים, לבחון נהלים, לזהות סטיות חוזרות ולחבר בין מופעים נפרדים של אותו כשל היא שמקנה לו יתרון מוסדי מובהק בבחינת מערכות AI ציבוריות. יתרון זה מתחדד בישראל גם נוכח תפקידו של מבקר המדינה כנציב תלונות הציבור, שכן בירור של תלונה פרטנית עשוי לחשוף ליקוי מערכתי רחב יותר הדורש תיקון כללי ולא מענה נקודתי בלבד. במובן זה, הפיקוח על שימוש ציבורי בבינה מלאכותית אינו נוגע רק לשאלת תקינותה של החלטה מסוימת, אלא גם, ואולי בעיקר, ליכולת לזהות בזמן כשל מערכתי בטרם יעמיק את פגיעתו בציבור רחב.

הפסיקה ברחבי העולם והספרות ההשוואתית כבר מספקות דוגמאות מוחשיות לכך שהסיכונים הכרוכים בשימוש במערכות אלגוריתמיות אינם תיאורטיים בלבד. אחת הדוגמאות הבולטות היא פרשת SyRI (System Risk Indication) בהולנד: רשויות הרווחה החלו להשתמש במערכת ממשלתית שנועדה לזהות סיכון להונאה באמצעות הצלבת מידע ממאגרי מידע ממשלתיים שונים, לרבות מידע שאספו גופים ציבוריים נוספים. בית המשפט בהאג קבע כי ההסדר הנורמטיבי שאפשר את השימוש במערכת אינו עומד בדרישות סעיף 8 לאמנה האירופית לזכויות אדם, בין היתר בשל הפגיעה הבלתי מידתית בפרטיות והיעדר שקיפות מספקת ביחס

Danielle Keats Citron, *Technological Due Process*, 85 Washington Univ. Law Rev. 1249, 1268 (2008). 6

Jenna Burrell, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, Big Data & Society 3 (2016). 7

מיעוט. פסק דין זה ממחיש כי גם כאשר בית משפט אינו פוסל את עצם השימוש בכלי אלגוריתמי, אין פירוש הדבר הכשרה מלאה ובלתי מסויגת שלו⁸.

בכך מבטא פסק הדין עמדה מורכבת יותר: לא של פסילה מוחלטת, אך גם לא של קבלה חסרת תנאים. דווקא המתח הזה, בין הסתייגות טכנולוגית לבין שימור שיקול דעת אנושי ואחריות משפטית, הוא שעומד בלב שאלת הפיקוח על מערכות AI ציבוריות. מן המקרים הללו עולה דפוס משותף: בכל אחד מהם, הדיון במערכת האלגוריתמית לא התמקד רק בדיוק הטכני שלה, אלא גם בשאלות רחבות יותר של שקיפות, זכויות, אחריות ומידת ההסתמכות הראויה על פלט ממכון. במובן זה, הדוגמאות אינן רק המחשה נקודתית לסיכונים, אלא אינדיקציה לכך שהשאלה המרכזית איננה אם המדינה רשאית להשתמש בכלים המבוססים על בינה מלאכותית, אלא באילו תנאים מוסדיים, דיוניים ונורמטיביים שימוש כזה יכול להיחשב לגיטימי.

מסגרת נורמטיבית: הדין הקיים ופערי הרגולציה

הדיון בשימוש בבינה מלאכותית בגופים ציבוריים בישראל אינו מתנהל במצב של ריק נורמטיבי מוחלט. בשנים האחרונות החלה להתעצב בישראל תפיסת מדיניות ממשלתית ביחס לאסדרת התחום, אשר אינה מבוססת, לפחות בשלב זה, על רגולציה רוחבית אחידה בסגנון האירופי, אלא על גישה סקטוריאלית, הדרגתית ומבוססת סיכון.

גישה זו מבקשת להטיל על כל רגולטור ענפי את האחריות למפות את השימושים במערכות בינה מלאכותית בתחום שעליו הוא מופקד, לזהות את הסיכונים המתעוררים ולבחון את המענים הרגולטוריים המתאימים להם⁹.

בכך מבקשת המדיניות הישראלית לאזן בין הרצון שלא לבלום חדשנות טכנולוגית באמצעות חקיקה רוחבית מוקדמת מדי, לבין ההכרה בכך ששימוש בבינה מלאכותית עלול לעורר סיכונים מהותיים לזכויות יסוד, לשוויון, לפרטיות, לאחריות ולשקיפות. נקודת המוצא המוסדית לגישה זו קיבלה ביטוי ברור במסמך "עקרונות, מדיניות

לאופן פעולת המערכת"⁸. במוקד הקושי עמדה העובדה שהמערכת פעלה על בסיס הצלבת מידע ממקורות ממשלתיים רבים, והפרט לא היה יכול לדעת בוודאות מהם הנתונים והשקלולים שהובילו לסימונו כמי שנמצא ב"סיכון". בכך המחיש המקרה כיצד שימוש שלטוני במערכת אלגוריתמית עלול לפגוע לא רק בפרטיות, אלא גם ביכולת להבין את בסיס ההחלטה ולהעמידה לביקורת אפקטיבית. המקרה ממחיש כי כאשר המדינה מפעילה מערכת אלגוריתמית בעלת השפעה ממשית על זכויות הפרט, אין די בתכלית מינהלית ראויה - נדרשים גם מנגנוני הגנה אפקטיביים, שקיפות מספקת ואיזון נורמטיבי הולם.

דוגמה שונה אך משלימה עולה מן המחלוקת האמריקאית סביב מערכת COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), כלי להערכת סיכון פלילי המשמש, בין היתר, בשלבי המעצר, השחרור טרם המשפט, הפיקוח והענישה, לצורך הערכת הסיכון להישנות עבירה או לאי-התייצבות. תחקיר ProPublica הצביע על כך שהמערכת יצרה פער מובהקים בשיעורי הטעות בין נאשמים שחורים לבין נאשמים לבנים. בפרט נטתה המערכת לסווג נאשמים שחורים כבעלי סיכון גבוה בשיעור גבוה לעומת כלל האוכלוסייה, גם כאשר בפועל לא שבו לבצע עבירות. הדיון המתודולוגי סביב ממצאים אלה עוד לא הסתיים, אך המקרה הפך לסמל מרכזי של החשש מפני הטיות מובנות במערכות חיזוי אלגוריתמיות, בעיקר כאשר הן פועלות על בסיס נתוני עבר המשקפים פערים חברתיים ומוסדיים קיימים⁹.

לצד זאת, בפסק דין מוכר שעסק בתחום, בעניין ערעורו של State v. Loomis על פסיקה שהסתמכה בין השאר על ציון שהופק ממערכת COMPAS קבע בית המשפט העליון של ויסקונסין כי השימוש בהערכת הסיכון של המערכת בנסיבות המקרה לא הפר את זכותו של הנאשם להליך הוגן, אך הדגיש כי אין להשתמש בציון כדי לקבוע אם הנאשם ייאסר או מה תהיה חומרת העונש, וכי ההסתמכות על הכלי חייבת להיות מלווה באזהרות ברורות בדבר אופיו הקנייני, הישענותו על נתונים קבוצתיים, היעדר תיקוף מספק לאוכלוסייה המקומית והחשש מהטיה כלפי קבוצות

8 Stichting Privacy First v. State of the Netherlands, ECLI:NL:RBDHA:2020:1878 (Dist. Ct. The Hague Feb. 5, 2020).

9 Jeff Larson et al., *How We Analyzed the COMPAS Recidivism Algorithm*, ProPublica (23.5.2016), <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>.

State v. Loomis, 881 N.W.2d 749, 769 (Wis. 2016).

11 משרד החדשנות, המדע והטכנולוגיה ומחלקת ייעוץ וחקיקה (משפט כלכלי) עקרונות, מדיניות, רגולציה ואתיקה בתחום הבינה המלאכותית 2023 https://www.gov.il/he/pages/ai_23, (2023).

רגולציה ואתיקה בתחום הבינה המלאכותית 2023", שפרסמו משרד החדשנות, המדע והטכנולוגיה ומחלקת ייעוץ וחקיקה (משפט כלכלי) במשרד המשפטים.

המסמך קובע במפורש כי בשלב זה יש לפעול במסגרת רגולציה ענפית ולא במסגרת רגולציה רוחבית, וכי על הרגולטורים לפעול, ככל האפשר, בהתאמה לנעשה במדינות מפותחות ובארגונים בין-לאומיים ובהתבסס על ניהול סיכונים והתאמת הרגולציה לרמת הסיכון¹². לצד זאת, המסמך ממליץ על שימוש בכלי רגולציה גמישים ומדורגים כגון עקרונות אתיים, תקינה, רגולציה עצמית, מודולריות, פיילוטס וניסויים מבוקרים וכן על פיתוח רגולציה בשיתוף מומחים, גורמים בתעשייה, אנשי אקדמיה והציבור.

ואכן, ישראל לא אימצה בשלב זה מודל של חוק רוחבי אחד בעניין הבינה המלאכותית, אלא מודל של רגולציה מבוזרת, סקטוריאלית ומתואמת. המסמך האמור אינו מסתפק בקביעה מוסדית בדבר אופי הרגולציה, אלא גם משרטט את ציר הסיכונים והעקרונות שאמורים להנחות את הרגולטורים בפעולתם.

בין האתגרים המנויים בו נזכרים, בין היתר, אפליה, מעורבות אנושית, הסברתיות, גילוי, אמינות, עמידות, אבטחה, בטיחות, אחריות ופרטיות. כן מודגש בו כי כל רגולטור נדרש לזיהוי ולמיפוי של השימושים בבינה מלאכותית בגופים מפוקחים ואחרים בתחום שעליו הוא אמון, של הסיכונים הכרוכים בכך ושל דרכי ההתמודדות האפשריות. חשיבות הדברים לענייננו כפולה: ראשית, הם מלמדים כי קיימת בישראל תפיסת מדיניות המזהה את הסיכונים הכרוכים בשימוש ב-AI; שנית, הם מלמדים כי המדינה עצמה בחרה, לפחות לעת הזו, להפקיד את עיקר מלאכת הרגולציה והפיקוח בידי הרגולטורים הענפיים.

גישה זו קיבלה ביטוי יישומי מאוחר יותר גם במהלכים ענפיים קונקרטיים. כך למשל, בדוח הסופי שפורסם בדצמבר 2025 בעניין השימוש בבינה מלאכותית בסקטור הפיננסי, הוצגה עבודתו של צוות בין-משרדי שכלל את רשות שוק ההון, הפיקוח על הבנקים, הרשות לניירות ערך, רשות התחרות, משרד המשפטים ומשרד האוצר¹³.

עצם הקמתו של צוות כזה, בהרכב בין-רגולטורי ובשביל סקטור מוגדר, ממחישה היטב כי הרגולציה הישראלית המתהווה מבוססת על תפיסה של טיפול ענפי מתואם, ולא על רגולציה רוחבית אחת החלה באופן אחיד על כל המשק.

הדוגמה הפיננסית אף מחזקת את ההבנה כי המדינה רואה ב-AI נושא המחייב מאמץ מערכתי, אך כזה שמתורגם בפועל לזירה הסקטוריאלית. עם זאת, העובדה שקיימת בישראל תפיסת מדיניות סקטוריאלית אינה מבטלת את קיומם של פערים רגולטוריים של ממש.

ההפך הוא הנכון. דווקא משום שהמדינה בחרה במודל מבוזר, שבו על כל רגולטור לעצב את המענה בתחום שעליו הוא מופקד, מתחדדת השאלה כיצד יובטח היקף ההגנה על זכויות באופן אחיד ואם קיימים מנגנוני פיקוח אפקטיביים שיכולים לזהות פערים, אי-עקביות או חוסר ממש בהסדרה. במילים אחרות, האתגר הישראלי אינו רק היעדרו של חוק רוחבי, אלא גם הסיכון שפיצול רגולטורי ייצור אי-אחידות, פערי הגנה או עיכוב בטיפול בתחומים שבהם הרגולטור הענפי טרם גיבש מסגרת מספקת. על רקע זה מתחדדת חשיבות ההסתמכות על עקרונות הדיון הישראלי הכללי, ובפרט דיני הפרטיות, חופש המידע והמשפט המינהלי. דינים אלה לא נועדו במקורם להסדיר שימוש בבינה מלאכותית, אך הם מספקים מסגרת נורמטיבית בסיסית החלה גם כאשר גוף ציבורי נשען על מערכות אלגוריתמיות.

חוק הגנת הפרטיות, התשמ"א-1981, מעגן את ההגנה על פרטיות ועל מידע אישי והוא נקודת מוצא מרכזית לכל בחינה של עיבוד מידע בידי רשויות¹⁴. עם זאת, החוק אינו כולל כיום הסדר מפורש בדבר החלטות המבוססות על עיבוד אוטומטי ואינו יוצר מנגנון דומה להסדר האירופי בתחום זה.

חוק חופש המידע, התשנ"ח-1998, עשוי לשמש בסיס עקרוני לדרישת גילוי גם ביחס לשימוש במערכות המשפיעות על הפרט. עם זאת, כאשר מתבקש גילוי מפורט יותר בדבר אופן פעולת המערכת, הנתונים שעליהם היא נשענת או מאפייני המודל שבו משתמשים, עשויות להתעורר טענות מצד הרשות או הספק בדבר סוד מסחרי, קניין רוחני או מגבלות אחרות על חשיפת המידע¹⁵. במקביל, עקרונות היסוד של המשפט המינהלי, ובהם חובת ההנמקה, החובה להפעיל שיקול דעת עצמאי, זכות הטיעון ואיסור השרירות, מוסיפים לחול גם כאשר הרשות נעזרת בכלים טכנולוגיים.

ואולם, הפסיקה הישראלית טרם פיתחה דוקטרינה ייעודית וסדורה ביחס להיקף ההסבר, הגילוי והבקרה הנדרשים כאשר החלטה מינהלית נשענת על מערכת AI.

12 ש.ם.

13 רשות שוק ההון, ביטוח וחיסכון בינה מלאכותית בסקטור הפיננסי, דוח סופי (2025).

14 חוק הגנת הפרטיות, התשמ"א-1981, ס"ח 128.

15 חוק חופש המידע, התשנ"ח-1998, ס"ח 227.

המסגרת המוצעת כאן שואבת השראה משלושה מקורות מרכזיים: מסגרת ניהול הסיכונים של המכון הלאומי לתקנים ולטכנולוגיה של ארצות הברית (National Institute of Standards and Technology; NIST) מחזורית ורב-שלבית לניהול סיכונים AI; הדין האירופי, ובפרט הוראות ה-EU AI Act ביחס למערכות בסיכון גבוה; והספרות והפסיקה שעמדו על הסכנות שבאטימות, בהטיה, בפיקוח פורמלי בלבד ובפגיעה בזכויות דיוניות.

מסגרת NIST AI RMF 1.0 מדגישה כי ניהול סיכונים השימוש בבינה מלאכותית אינו פעולה חד-פעמית אלא תהליך מתמשך הנשען על ארבע פונקציות ליבה: ממשל ברור, מיפוי שימושים וסיכונים, מדידה והערכה של השפעות וניהול שוטף של הסיכון לאורך חיי המערכת (Govern, Map, Measure, Manage)¹⁸. מסגרת זו אומנם אינה מחייבת משפטית בישראל, אך היא מספקת היגיון ארגוני שימושי במיוחד לביקורת מוסדית.

ה-EU AI Act דורש כי בעת שימוש במערכות בסיכון גבוה יקפידו על תיעוד, ניהול סיכונים, לוגים, פיקוח אנושי, דיוק, עמידות ואבטחה לאורך כל מחזור החיים של המערכת.

להלן שישה עקרונות מוצעים לבחינה מוסדית. אין לראות בהם רשימת בדיקה טכנית בלבד, אלא מערכת משולבת של תנאים הכרחיים לשימוש אחראי בבינה מלאכותית במרחב הציבורי.

עיקרון 1: שקיפות ותיעוד

נקודת המוצא לביקורת אפקטיבית היא קיומו של תיעוד. גוף ציבורי אינו יכול לעשות שימוש אחראי במערכת בינה מלאכותית אם אינו מחזיק בתיעוד מספק של מאפייניה, ייעודה, מגבלותיה, מקורות הנתונים שעליהם היא נשענת, זהות הספק, גרסת המודל, אופן העדכון, תוצאות הביצועים הידועות והגורם המוסדי האחראי להפעלה.

בהיעדר תיעוד כזה, לא ניתן לקיים פיקוח אמיתי, לא ניתן לעקוב אחר שינויים בגרסאות ולא ניתן לייחס אחריות כאשר מתגלה כשל. למעשה נדרש Model Card - מסמך תיאורי שיטתי של המערכת. אין הכרח לאמץ דווקא את המונח הזה, אך חשוב לאמץ את ההיגיון: כל מערכת צריכה להיות ניתנת לזיהוי, להסבר מוסדי בסיסי ולהערכה בדיעבד. על כן, יש לשאול לא רק אם המערכת קיימת, אלא אם קיימת עבורה תשתית תיעודית נאותה.

לשם השוואה, הדין האירופי מציע כיום מסגרת מפורטת ומתקדמת יותר. הרגולציה האירופית להגנת המידע (General Data Protection Regulation; GDPR) כוללת הגנות ביחס להחלטות המבוססות אך ורק על עיבוד אוטומטי, לצד חובות שקיפות ואמצעי הגנה דיוניים¹⁶, ואילו חוק הבינה המלאכותית של האיחוד האירופי (EU AI Act) מאמץ משטר רגולטורי מבוסס סיכון הקובע דרישות מחמירות ביחס לשימוש בבינה מלאכותית בתחומים מסוימים המוגדרים כבעלי סיכון גבוה, כגון תעסוקה, חינוך, שירותים ציבוריים חיוניים, אכיפת חוק, הגירה וניהול הליכים שיפוטיים או מינהליים¹⁷.

ההשוואה לדין האירופי לא נועדה בהכרח להכתיב אימץ מלא של המודל האירופי בישראל, אך היא ממחישה עד כמה ההסדרה הישראלית נותרה בשלב זה חלקית, מבזרת ותלויה ביכולתם של רגולטורים ענפיים לפתח מענה עצמאי. מכאן נגזרת גם חשיבותה האפשרית של ביקורת המדינה. במצב שבו הרגולציה הישראלית מבוססת על חלוקה סקטוריאלית, על עקרונות מדיניות כלליים ועל דינים כלליים שאינם ייעודיים לשימוש בבינה מלאכותית, גובר הצורך בגוף ביקורת בעל ראייה מערכתית, המסוגל לבחון לא רק אם כל רגולטור פועל בתחומו, אלא גם אם המבנה הכולל יוצר הגנה מספקת, עקבית ואפקטיבית על הציבור. תפקידה של ביקורת המדינה בהקשר זה אינו להחליף את הרגולטורים הענפיים, אלא לבדוק אם המדיניות שהותוותה אכן מתורגמת הלכה למעשה למנגנוני פיקוח, תיעוד, בקרה והגנה על זכויות.

שישה עקרונות לבחינה מוסדית

הפרק הנוכחי מבקש לעבור מן המישור הנורמטיבי אל המישור היישומי. הפרקים הקודמים עסקו בשאלה מדוע השימוש הציבורי בבינה מלאכותית מעורר קושי משפטי ומוסדי, ופרק זה ידון בשאלה מה צריך מבקר המדינה לבדוק כאשר גוף ציבורי משתמש במערכת AI. התשובה המוצעת אינה נשענת על עיקרון אחד בודד, אלא על מסגרת מוסדית משולבת, המיועדת לבחינת מערכות בינה מלאכותית לאורך מחזור חייהן המלא - משלב הרכש וההטמעה, דרך שלב ההפעלה השוטפת ועד לשלב העדכון, ההשעיה או הפסקת השימוש.

Regulation (EU) 2016/679 of the European Parliament and of the Council (General Data Protection Regulation), arts. 13, 14, 15 & 22, OJ L 119/1 (4.5.2016).

Regulation (EU) 2024/1689 L, לעיל ה"ש 4, סעיף 6.

National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (2023), chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/<https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf>.

המשפט הקיים כבר מכיר אומנם בצורך בכלים כמו DPIA (Data Protection Impact Assessment) בעת עיבוד הכרוך בסיכון גבוה, ויש לכך התייחסות ב-GDPR, אך הכוונה היא בעיקר למצבים שבהם העיבוד עשוי להשפיע על זכויות וחירויות, ובפרט כאשר מדובר בהערכה שיטתית ואוטומטית של מאפיינים אישיים, בקבלת החלטות בעלות השפעה ניכרת או בעיבוד רחב היקף של מידע רגיש. הערכת ההשפעה של AI במגזר הציבורי צריכה להיות רחבה יותר²¹. יש לשאול מי עלול להיפגע, אילו קבוצות מושפעות במיוחד, האם קיימת חלופה פחות פוגענית, מהי רמת ההסתמכות המתוכננת על המערכת, מהן ההשלכות של כשל אפשרי, מי אישר את השימוש במערכת ואיזו תשתית תיעודית היא תכלול.

עיקרון 4: אחריותיות ומנגנוני השגה

החלטה מינהלית הנשענת על מערכת אלגוריתמית אינה חדלה להיות החלטה מינהלית. משום כך, עליה לעמוד בדרישות היסוד של אחריותיות: יש לוודא כי יהיה ניתן להבין מה עמד בבסיסה, לזהות מי נושא באחריות לה ולאפשר לפרט שנפגע ממנה לבקש בחינה חוזרת או להשיג השגות. אם ההחלטה מוצגת כקביעה של מחשב, ההנמקה נפגעת ושרשרת האחריות מתערפלת.

לעיתים הרשות מפנה לספק, הספק מפנה למודל, והפרט נותר ללא כתובת ברורה. דווקא כאן יש חשיבות מיוחדת לביקורת מוסדית. מבקר המדינה צריך לבחון אם קיימים מסלולי השגה ברורים, אם יש גורם אנושי מוסמך האחראי להחלטה, אם ניתן לשחזר את הנתונים והשלבים שהובילו לתוצאה ואם האחריות אינה מתפזרת על פני שרשרת האספקה באופן המסכל ביקורת. דרישה זו נובעת הן מעקרונות המשפט המינהלי והן מהדיון הרחב יותר על אחריותיות בשרשראות טכנולוגיות מורכבות.

עיקרון 5: הכחתת הטיה ובחינת נתוני אימון

העובדה שמערכת AI אינה "רואה" משתנה של גזע, מגדר או שיוך קבוצתי באופן מפורש, אינה מבטיחה היעדר אפליה. זהו אולי אחד הלקחים המרכזיים שעלו מן הספרות ומהמקרים האמפיריים הבולטים. מערכות חיצוניות ולמידת מכונה עלולות לשעתק דפוסיים היסטוריים של אי-שוויון גם כאשר הן נשענות על משתנים הנחזים כניטרליים. משום כך, בחינת הטיה אינה יכולה להסתפק בהצהרה שהאלגוריתם "עיוור" למאפיין מוגן. נדרשת בדיקה שיטתית

הוראות ה-EU AI Act כוללות דרישות דומות לתיעוד טכני, למידע למפעילים, ללוגים ולשמירת תיעוד עבור הרשויות המוסמכות¹⁹. עם זאת, לצורך ביקורת מדינה אין הכרח להמתין להסדר ישראלי זהה. די להבין שתיעוד הוא תנאי ראשוני לכל פיקוח. מובן כי תיתכן שאלת ביקורת מעשית: האם לכל מערכת AI פעילה בגוף המבוקר יש מסמך תיאורי עדכני הכולל את מטרת השימוש, הגרסה הפעילה, זהות הספק, סוגי הנתונים הרלוונטיים, ביצועים ידועים, מגבלות ידועות והגורם האחראי בארגון.

עיקרון 2: בקרה אנושית ממשית

המונח "פיקוח אנושי" (Human oversight) הפך לביטוי כמעט שגרתי בכל מסמך רגולטורי העוסק בבנייה מלאכותית, אך לא כל פיקוח אנושי הוא פיקוח ממש. לא פעם, תהליך קבלת ההחלטות מציג כלפי חוץ מעורבות של פקיד או בעל סמכות, אך בפועל האדם רק מאשר את פלט המערכת בלי להבין לעומק את משמעותו, בלי לבחון אפשרות לסטות ממנו ובלי לתעד שיקול דעת עצמאי.

במצב כזה, הנוכחות האנושית היא פורמלית בלבד, ואין בה כדי לספק את ההגנה המוסדית הנדרשת. סיתרון (Citron) עמדה כבר לפני שנים על הסכנה שבמנגנוני procedural safeguards הנחצים כלפי חוץ כהגנה דיונית מספקת, אך בפועל אינם אלא טקס ריק של מקבלי החלטות הנוטים לייחס למערכת הממוחשבת אמינות יתר או שאינם מצוידים בכלים המאפשרים להבין את הפלט שלה²⁰.

ה-EU AI Act מתייחס לכך במפורש ודורש שתכנון של מערכות בסיכון גבוה יאפשר פיקוח אנושי אפקטיבי, שמטרתו לצמצם או למנוע סיכונים לבריאות, לבטיחות ולזכויות יסוד. המשמעות בעניין ביקורת המדינה ברורה: אין די בכך שהחלטה "עוברת דרך אדם". יש לבחון אם האדם באמת מסוגל להפעיל שיקול דעת עצמאי, להבין את ההמלצה, להטיל בה ספק ולחרוג ממנה בעת הצורך.

עיקרון 3: הערכת השפעת המערכת לפני הטמעתה

אחד הכשלים המרכזיים בהטמעת מערכות AI במגזר הציבורי הוא המעבר המהיר מדי משלב ההתלהבות הטכנולוגית לשלב ההפעלה בפועל, בלי שנערכה בחינה מוקדמת ומסודרת של הסיכונים. כאן נדרשת הערכה של השפעת המערכת טרם הפעלתה. יש להעריך את ההשפעה על הפרטיות וכן על זכויות, שוויון, שקיפות, תלות טכנולוגית, אחריותיות וסיכונים הטיית.

19 Regulation (EU) 2024/1689, לעיל ה"ש 4, סעיפים 9 - 15.

20 Citron, Technological Due Process 20, לעיל ה"ש 6, 149, 1287 - 1295.

21 Regulation (EU) 2016/679, לעיל ה"ש 16, סעיף 35.



בתוך מערכות ציבוריות שכבר פעלו בתנאים של אי-שוויון או נגישות לא אחידה לשירותים.

תחקיר ProPublica על COMPAS ממחיש זאת בבירור²². אף שאין להתייחס אליו כאל פסק דין או אמת מדעית סופית, התחקיר מראה כיצד פערים בשיעורי הטעות בין קבוצות אוכלוסייה שונות עשויים להופיע גם במערכת שמתיימרת להיות מכשיר ניטרלי.

של ייצוג הנתונים, איכותם, מקורותיהם וההשפעה בפועל של המערכת על קבוצות שונות.

בכלל זה יש לבחון אם נתוני האימון משקפים באופן מאוזן את האוכלוסייה הרלוונטית, אם הם כוללים קבוצות שאינן מיוצגות די הצורך, ואם הם נושאים עימם הטיות היסטוריות או מוסדיות שהמערכת עלולה לשעתק. בחינה זו חשובה במיוחד כאשר המודל אומן על בסיס נתוני עבר שנוצרו

22 Julia Angwin et al., *Machine Bias*, ProPublica (23.5.2016), <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

תפקיד מבקר המדינה: לקראת ביקורת מותאמת AI

ששת העקרונות שהוצגו בפרק הקודם אינם יכולים להישאר בגדר מסגרת נורמטיבית מופשטת בלבד. כדי שיהפכו לאמות מידה אפקטיביות, עליהם להיות מתורגמים לכלים מוסדיים, מתודולוגיים וארגוניים המותאמים לאופי הייחודי של מערכות בינה מלאכותית.

כאן בדיוק מתחדדת שאלת תפקידו של מבקר המדינה, אשר בוחן חוקיות, תקינות, יעילות וחיסכון במובנם המסורתי, אך נדרש להתאים את ארגז הכלים הביקורתי שלו למציאות שבה החלטות ציבוריות מושפעות ממודלים, נתונים, ספקים חיצוניים, מערכות לומדות ותשתיות דיגיטליות מורכבות.

סמכותו הרחבה של מוסד מבקר המדינה ואחריותו כלפי הכנסת ולא כלפי הממשלה מציבות אותו בעמדה מוסדית ייחודית לבחינה מסוג זה²³. עם זאת, עצם קיומה של סמכות רחבה אינו מבטיח כשלעצמו יכולת אפקטיבית לפעול בתחום חדש זה. במובן זה, האתגר אינו נובע מהיעדר סמכות פורמלית, אלא מהפער שבין כלי הביקורת המסורתיים לבין האובייקט החדש שיש לבדוק.

המבקר יודע היטב לשאול אם הליך רכש בוצע כדון, אם נשמרה הפרדה בין סמכויות, אם קיימים נהלים פנימיים ואם הוקצו תקציבים בהתאם להחלטות המוסמכות. אך כאשר מדובר במערכת AI, שאלות הביקורת צריכות להתרחב באופן ניכר: האם בוצעה הערכת סיכון טרם ההטמעה, האם נבחנו פערי הטיה בין קבוצות, האם קיימת בקרה אנושית ממשיית שאינה אישור פורמלי בלבד, האם נשמרים לוגים ותיעוד טכני, האם קיימת תלות בספק יחיד והאם קיימים מסלולי השגה אפקטיביים לאזרח שנפגע מהחלטה הנתמכת במערכת.

מסגרות כמו NIST AI RMF 1.0, המדגישות ממשל, מיפוי, מדידה וניהול מתמשך של סיכונים AI, ממחישות היטב את הצורך לעבור מביקורת נקודתית לביקורת מחזורית ומבוססת סיכון²⁴. דוח של משרד הביקורת הלאומי של בריטניה מ-2024 על השימוש ב-AI בממשלה מחזק אף הוא את ההבנה שהצלחה או כישלון בתחום זה תלויים לא רק בטכנולוגיה עצמה, אלא גם בשאלות כגון מי אמון על המערכת בארגון, כיצד נבדקת איכות הנתונים שעליהם היא נשענת, האם לעובדים המשתמשים בה יש ידע והכשרה מספקים והאם מתקיים תיאום בין הגורמים המינהליים,

מבקר המדינה אינו צריך להפוך למעבדת מחקר אלגוריתמי, אך עליו לדרוש מן הגוף המבוקר להראות אם נבדקה הטיה, באילו שיטות, באיזו תדירות ומה נעשה כאשר מתגלים פערים.

עיקרון 6: אבטחה, עמידות והיערכות לכשל

מערכות AI ציבוריות חשופות לשאלות של הוגנות ושקיפות, אמינות, אבטחה, עמידות ותפקוד תחת תנאי שינוי. AI Act דורש ממערכות בסיכון גבוה רמה מתאימה של דיוק, חוסן ואבטחת סייבר, ומסגרת NIST מדגישה כי ניהול הסיכונים צריך להיות רציף ומחזורי וכי עליו להתעדכן לאורך חיי המערכת.

על מוסדות המשתמשים במערכות AI לבחון לא רק אם המערכת פעלה היטב ביום ההשקה, אלא גם כיצד היא מתמודדת עם סטיית מודל (drift), עם שינויים באוכלוסיית המשתמשים, עם עדכוני תוכנה, עם תקיפות סייבר או עם כשל של ספק. בהקשר זה עולה גם שאלת התלות בספק (vendor lock-in).

אם גוף ציבורי רוכש מערכת שאינו יכול לבדוק, שאינו יכול להחליף בקלות או שאינו מחזיק בגישה מספקת לנתונים, ללוגים שלה או לתיעוד הרלוונטי, הוא עלול למצוא עצמו תלוי בספק באופן הפוגע בביקורת, בבקרה ואף ברציפות השירות.

על פי ה-AI Act, בעת שימוש במערכות AI המוגדרות כמערכות בסיכון גבוה, קיימות דרישות ברורות ביחס ללוגים ולתיעוד מסמכים ושמירתם. גם ללא העתקה ישירה למשפט הישראלי, ניתן לגזור מה-AI Act אמת מידה מוסדית חשובה: שימוש אחראי ב-AI מחייב היערכות לכשל, לגיבוי תהליכים אנושיים ולניהול תלות בספק באופן שאינו מסכל ביקורת.

ששת העקרונות אינם רשימה אקראית של דרישות נפרדות. הם שלובים זה בזה: תיעוד הוא תנאי לפיקוח אנושי אפקטיבי, פיקוח אנושי הוא תנאי לאחריותיות, אחריותיות תלויה באפשרות להשיג על החלטה, בחינת הטיה תלויה באיכות התיעוד והמדידה ואבטחה ועמידות משלימות את כל אלה בכך שהן מוודאות שהמערכת אינה רק "הוגנת" בתיאוריה, אלא גם בשלטה, נבדקת ומבוקרת בפועל לאורך זמן. ביקורת מדינה המבקשת להתמודד ברצינות עם AI ציבורי תצטרך אפוא לבחון את מכלול העקרונות יחד ולא להסתפק בקיום פורמלי של אחד מהם.

23 חוק-יסוד: מבקר המדינה, סעיפים 2 ו-6.

24 NIST, *NIST; 18*, לעיל ה"ש NIST, *Artificial Intelligence Risk AI RMF Playbook* (2023) <https://www.nist.gov/itl/ai-risk-management-framework/nist-ai-rmf-playbook>.

המשפטיים והטכנולוגיים המעורבים בהפעלתה²⁵.

על רקע זה, ניתן להציע חמש המלצות מוסדיות מרכזיות:

ההמלצה הראשונה היא יצירת מסגרת בדיקה שיטתית ומובנית שתאפשר להפוך את עקרונות הביקורת על AI לכלי עבודה מעשי. נכון שמבקר המדינה יפתח שאלון ביקורת ייעודי לשימוש בביקורות של גופים ציבוריים המשתמשים במערכות AI.

שאלון כזה לא צריך להיות מסמך טכני סגור וקשיח, אלא מסגרת מודולרית שתתאים את עצמה לאופי השימוש. נדרשת למשל הבחנה בין מערכות תומכות החלטה פנימיות, מערכות דירוג סיכון, מערכות המכוונות הקצאת משאבים או מערכות המתקשרות ישירות עם הציבור.

חשיבותו של שאלון כזה כפולה: ראשית, הוא מספק למבקרים שפה משותפת וסדרה עקבית של שאלות, ובכך מצמצם את הסיכון לביקורת חלקית, מקרית או אינטואיטיבית מדי. שנית, פרסומו עשוי לייצר אפקט מניעתי חשוב. גופים ציבוריים ידעו מראש מה צפוי להיבחן ויוכלו להיערך בהתאם באמצעות תיעוד מסודר, ניהול סיכונים והטמעת בקרה פנימית טובה יותר.

במובן זה, השאלון אינו רק מכשיר ביקורת בדיעבד, אלא גם כלי ממשל מקדים. מסגרות כמו NIST AI RMF Playbook וממצאי ה-NAO בדבר הצורך בתפקידי בעלות, אחריות ותיאום בממשלה מספקות בסיס טוב לפיתוח כלי כזה, גם אם נדרשת התאמה שלהן להקשר הישראלי.

ההמלצה השנייה נוגעת לפיתוח מומחיות טכנית פנימית. ביקורת אפקטיבית על מערכות AI מחייבת הבנה טכנולוגית מספקת. אין הכוונה לכך שכל מנהל ביקורת יהפוך למדען נתונים או למהנדס למידת מכונה, אלא לכך שמוסד הביקורת יפתח מוקד מומחיות פנימי - יחידה ייעודית או צוות רוחבי תומך - שיסייע לצוותי הביקורת הסקטוריאליים בבדיקות הנוגעות לאלגוריתמיקה, נתונים, מודלים, הטיה, אבטחה ותיעוד טכני.

ה-NAO מדגים כי מוסד ביקורת המדינה יכול להישען על תשתיות ניתוח נתונים ושירותי דאטה כדי לייעל עבודת ביקורת, לחבר מאגרי מידע, לזהות דפוסים ולבסס ממצאים באופן אמפירי יותר²⁶. גם אם לא מדובר בדיוק ביחידת AI ייעודית במובן המצומצם, המודל הבריטי בהחלט תומך בהיגיון של שילוב יכולות אנליטיות וטכנולוגיות בעבודת

הביקורת. המשמעות היישומית לישראל היא ברורה. בהיעדר מומחיות טכנולוגית מספקת, ביקורת AI עלולה להצטמצם לבדיקה פורמלית של נהלים, בלי לחדור אל לב הבעיה. גוף מבוקר עשוי להציג "מדיניות AI", "תיעוד", "אישור אנשי" או "בדיקות תקופתיות", אך רק צוות בעל הבנה מספקת יוכל לבחון אם מדובר במנגנון ממשי או בתחליף סמלי.

השקעה ביחידה מומחית כזו עשויה להיות יקרה בטווח הקצר, אך תשואתה הציבורית עשויה להיות גבוהה מאוד, משום שכשל מערכתי אחד במערכת AI ציבורית עשוי להשפיע על אוכלוסייה רחבה ביותר.

ההמלצה השלישית נוגעת לצורך בשיתוף פעולה מוסדי. השימוש הציבורי בבניה מלאכותית יושב על התכר שבין כמה תחומים רגולטוריים: פרטיות, הגנת מידע, שוויון, הליך מינהלי תקין, רגולציה ענפית, אבטחת מידע ועוד. משום כך, ביקורת המדינה אינה יכולה ואינה צריכה לפעול כאילו מדובר בטריטוריה מוסדית השייכת לה בלבד. בפרט, קיימת חפיפה מסוימת, גם אם לא זהות, בין עולמה של הרשות להגנת הפרטיות לבין תחום הביקורת על שימוש ציבורי ב-AI.

הרשות עוסקת בעיבוד מידע אישי ובהיבטי פרטיות, ואילו מבקר המדינה בוחן תקינות מוסדית רחבה יותר. דווקא בשל כך, קיים ערך משמעותי לשיתופי פעולה ממוסדים, לפחות ברמת החלפת ידע, מתודולוגיות, מוקדי סיכון ודפוסי אכיפה.

תפקידו של מבקר המדינה בהקשר זה אינו לבטל את המבנה הענפי, אלא לבחון אם הוא אכן פועל, אם קיימת בו עקביות ואם נוצרו פערים בין רגולטורים שונים או בין תחומים שונים של השירות הציבורי.

ההמלצה הרביעית עניינה לעודד היערכות מוקדמת של גופים ציבוריים באמצעות פרסום מסמך מנחה פומבי. הביקורת על AI אינה צריכה להתחיל רק לאחר שכבר נגרם נזק. לצד שאלון הביקורת הפנימי של מבקר המדינה, ניתן לגבש מסמך מנחה שאינו חקיקה ואינו רגולציה מחייבת במובן הפורמלי, אך יוכל לסייע לגופים ציבוריים להבין מראש מהן הציפיות הבסיסיות ביחס לשימוש אחראי בבניה מלאכותית.

כשם שגופים ממשלתיים מסתמכים על מדריכים לניהול סיכונים, לבקרה פנימית, לממשל תאגידי או לניהול מידע,

UK National Audit Office, *Use of Artificial Intelligence in Government* (2024), <https://www.nao.org.uk/reports/use-of-artificial-intelligence-in-government/>.

National Audit Office, *How Data Analytics Can Help with Audits* (2020), <https://www.nao.org.uk/insights/how-data-analytics-can-help-with-audits/>; National Audit Office, *Using Data Analytics to Tackle Fraud and Error* (2025), <https://www.nao.org.uk/reports/using-data-analytics-to-tackle-fraud-and-error/>.

סיכום

המסגרת שהוצעה במאמר זה אינה מבקשת להרתיע גופים ציבוריים מפני שימוש בבינה מלאכותית ואינה נשענת על עמדה ריאקציונית שלפיה עצם הטמעת הטכנולוגיה היא בגדר תקלה מוסדית. ההפך הוא הנכון. במקרים רבים, מערכות בינה מלאכותית עשויות לשפר את איכות השירות הציבורי, לייעל תהליכים אדמיניסטרטיביים, לצמצם זמני טיפול, לסייע בזיהוי צרכים שלא זוהו קודם לכן ולאפשר הקצאה מדויקת יותר של משאבים ציבוריים מוגבלים. דווקא משום שהפוטנציאל הציבורי של טכנולוגיות אלה הוא ממש, השאלה המרכזית אינה אם להתיר את השימוש בהן, אלא באילו תנאים משפטיים, מוסדיים וביקורתיים ניתן להבטיח שהשימוש יהיה אחראי, הוגן ונתון לפיקוח אפקטיבי.

מנקודת מבט זו, הדיון בבינה מלאכותית בשירות הציבורי אינו דיון טכנולוגי גרידא. הוא נוגע בליבה של תפיסת השלטון החוקתי והמינהלי. כאשר החלטה על זכאות, על גישה לשירות, על הקצאת משאב, על סימון "סיכון" או אף על התערבות עמוקה בחיי הפרט נשענת על פלט אלגוריתמי, אין מדובר רק בשאלה של יעילות, אלא בשאלה של כוח שלטוני: מי מקבל את ההחלטה, על סמך אילו נתונים, באיזו מידה ניתן להבין אותה, מי נושא באחריות לה וכיצד יכול הפרט להתמודד עימה.

במובן זה, השאלה המרכזית אינה אם האלגוריתם "חכם" דיו, אלא אם מוסדות המדינה נותרו נאמנים לחובות היסוד שלהם גם כאשר הם פועלים באמצעות מערכות חישוביות מורכבות. חזרה אל דוגמת הרווחה הפותחת ממחישה את הדברים היטב. אם בוחנים את המקרה דרך פריזמת ששת העקרונות שהוצעו במאמר, מתגלה רצף של כשלים מצטברים: היעדר תיעוד מספק של המערכת, היעדר בקרה אנושית ממשית עליה, היעדר בחינה מוקדמת של השפעותיה על זכויות ועל קבוצות כגיעות, היעדר מנגנון ברור להשגה או לערר, היעדר בחינה שיטתית של הטיה בתוצאות והיעדר בחינה של תלות בספק ושל היערכות לכשל. רצף כזה אינו רק כשל טכנולוגי או ניהולי, הוא כשל מוסדי, וכאשר כשל מוסדי כזה אינו מזוהה, אינו נבדק ואינו מתוקן בזמן, הוא עלול להפוך במהירות לכשל חוקתי-מינהלי המשפיע על ציבור רחב.

כאן בדיוק מצוי תפקידו האפשרי, ואף המתבקש, של מבקר המדינה. מבקר אינו אמור להפוך לרגולטור AI במובן המלא, ואינו מחליף את המחוקק, את הממשלה או את הרגולטורים הענפיים. תפקידו שונה: לחשוף פערים, לזהות כשלים מבניים, להצביע על היעדר מנגנוני הגנה ולהפוך בעיות

כך ניתן לגבש מדרך עקרונות לשימוש אחראי ב-AI בגופים ציבוריים²⁷. מדרך כזה לא יצטרך להכריע בכל שאלה משפטית מורכבת, אך כן יוכל לנסח ציפיות ברורות ביחס לתיעוד, להערכת סיכון, לבקרה אנושית, לשקיפות, לניהול ספקים, לניטור הטיית ולהיערכות לכשל.

הרעיון הזה נתמך גם בניסיון הממשלתי הרחב שנצבר בישראל בתחום ניהול הסיכונים ברגולציה ובמדיניות ציבורית. מדריכים רוחביים אינם יוצרים לבדם משטר משפטי מלא, אך הם מסייעים ביצירת אחידות מושגית, שפה מוסדית משותפת וסטנדרט מקצועי שהמבוקר והמבקר גם יחד יכולים להתייחס אליו. לכן, מדרך כזה יכול לשמש בעת ובעונה אחת גם ככלי הכוונה מוקדם לגופים ציבוריים וגם כאמת מידה ביקורתית בבדיקות עתידיות.

ההמלצה החמישית היא שילוב של למידה ושיתוף פעולה בין-לאומיים, אך בזהירות מושגית. ביקורת הבודקת שימוש בבינה מלאכותית יכולה להסתייע הן בדוחות קונקרטיים של מוסדות ביקורת עליונים והן בעקרונות הבין-לאומיים הרחבים של ניהול סיכונים, בקרה פנימית ואחריות ציבורית. אין להתעלם מן הערך של למידה ממוסדות ביקורת עליונים וממסגרות מקצועיות בין-לאומיות: ה-NAO הבריטי כבר עסק ישירות בשימוש ב-AI בממשלה; מוסדות ביקורת אחרים בעולם מפתחים בהדרגה כלים אנליטיים, יכולות של ניתוח נתונים (data analytics) ומסגרות ביקורת מותאמות לעולם דיגיטלי; ו-INTOSAI בכללותה מדגישה זה שנים את הקשר בין בקרה פנימית, ניהול סיכונים, ואחריות וממשל תקין.

עם זאת, כאן בדיוק דרושה זהירות: לא כל מסמך בין-לאומי שמדבר על ממשל תקין (good governance) או ניהול סיכונים (risk management) הוא מסמך ייעודי העוסק ב-AI, ולא נכון "לצבוע" בדיעבד מסמכים כלליים כסטנדרטים ייעודיים לביקורת בתחום.

המסקנה העולה מן האמור היא שהמעבר לביקורת מותאמת AI אינו מצריך מהפכה מוסדית מוחלטת, אך הוא בהחלט מחייב שינוי משמעותי של הכלים, הכשירויות והמתודולוגיה. סמכות מבקר המדינה לבדה אינה מספיקה. נדרשת גם בנייה מודעת של תשתית ביקורתית חדשה, הכוללת שפה מקצועית עדכנית, מומחיות טכנולוגית, שיתוף פעולה בין-מוסדי, כלי הנחיה מוקדמים ויכולת ללמוד מן הנעשה בעולם בלי לאמץ בצורה לא ביקורתית מונחים או מסמכים שאינם מתאימים במדויק למציאות הישראלית.

27 משרד ראש הממשלה, **מדרך לניהול סיכונים ברגולציה ובמדיניות ציבורית** (2018), [chrome-extension://efaidnbmnnnibpcjpcglclefindmkaj/ https://www.gov.il/BlobFolder/generalpage/reg-risk-management-guide/he/file_file_Guide171018.pdf](https://www.gov.il/BlobFolder/generalpage/reg-risk-management-guide/he/file_file_Guide171018.pdf).

על כך שהסכנה המרכזית אינה עצם קיומם של מודלים מורכבים, אלא האפשרות שכוח ניהולי, מסווג, מדרג ומכריע יופעל מאחורי מעטה של מורכבות טכנית, סודיות מסחרית או שמכות מקצועית לכאורה.

לפי יובנקס (Eubanks), הביקורת מתמקדת באופן שבו מערכות דיגיטליות מציגות הכרעות פוליטיות ומבניות כשאלות של יעילות אדמיניסטרטיבית, וכך נוצרים שחיקה של הליך הוגן (due process), שיעתוק אי-שוויון והעמקת פערי כוח²⁸. על פי פסקוואלה (Pasquale), הדגש הוא על הסכנה של "קופסה שחורה", כלומר מציאות שבה מוסדות חזקים בוחנים, מדרגים ומעריכים אחרים בלי להיחשף בעצמם לביקורת שקולה²⁹.

שתי התובנות האלה רלוונטיות במיוחד למרחב הציבורי. הן מלמדות כי לצד הבדיקה אם הטכנולוגיה מדויקת, חשוב לבדוק אם הכוח שהיא מפעילה נותר כפוף לביקורת, לחשיפה ולאחריותיות. בסופו של דבר, שאלת הבינה המלאכותית בשירות הציבורי אינה שאלה של חדשנות בלבד ואף לא שאלה של פרטיות בלבד. זו שאלה של ארכיטקטורת כוח.

כאשר רשות ציבורית נעזרת במערכת אלגוריתמית כדי לדעת יותר, להחליט מהר יותר ולפעול בהיקף רחב יותר, מתעצמת גם החובה להצדיק, להסביר, לבקר ולתקן. חברה דמוקרטית אינה נבחנת רק ביכולתה לאמץ טכנולוגיה מתקדמת, אלא גם ביכולתה להבטיח שהטכנולוגיה לא תשחרר את השלטון מן החובות הבסיסיות המוטלות עליו כלפי האזרח. אם בעבר ביקורת המדינה נועדה להגן על הציבור מפני שרירותיות שלטונית גלויה, הרי שבעידן הנוכחי עליה להסתגל גם לצורות חדשות, שקטות ומתוחכמות יותר של שרירותיות, הפועלות באמצעות נתונים, מודלים וציון סיכון.

לכן, הצעדים שהוצעו במאמר זה הם אינם "מהפכה" אלא תנאי סביר לשמירה על יסודות הממשל התקין בעידן אלגוריתמי, ומפני שהם צעדים סבירים קשה להצדיק את דחייתם. בעידן שבו קצב ההטמעה של מערכות AI מהיר בהרבה מקצב התגבשותם של כללי בקרה, האיפוק הביקורתי אינו ניטרליות אלא ויתור בפועל על יכולת ההגנה של המשפט הציבורי. אם מבקר המדינה יבקש למלא את תפקידו ההיסטורי גם במרחב הדיגיטלי, לא די שיבחן את תוצאות המדיניות, עליו לבחון גם את התשתיות החישוביות שמעצבות אותה.

טכנולוגיות שנוטות להישאר מוסתרות בתוך שיח מקצועי או קבלני לבעיות ציבוריות ומשפטיות גלויות. במציאות של רגולציה סקטוריאלית, מבזרת ומתפתחת, חשיבותו של גוף ביקורת בעל ראייה רוחבית אף גוברת, משום שהוא עשוי לבחון לא רק את התנהלותו של כל רגולטור בנפרד, אלא גם אם המכלול אכן יוצר הגנה מספקת, עקבית ואפקטיבית על הציבור.

במובן זה, התרומה האפשרית של ביקורת המדינה חורגת משאלת האכיפה במובנה הצר. ביקורת טובה אינה רק מצביעה על הפרה - היא גם מייצרת שפה מוסדית. אם מוסד הביקורת יבחן באופן עקבי תיעוד, בקרה אנושית, הערכת השפעה מוקדמת, מנגנוני השגה, בדיקות הטיה, אבטחה, עמידות ותלות בספקים, הוא יתרום לא רק לחשיפת ליקויים, אלא גם לעיצוב סטנדרט ציבורי חדש של שימוש אחראי בבינה מלאכותית. בכך, הביקורת עשויה למלא תפקיד כפול: תפקיד רטרופקטיבי של איתור מחדלים ותפקיד פרוספקטיבי של הכוונת התנהגות מוסדית עתידית.

מן העבר השני, אין לייפות את התמונה. ביקורת המדינה לבדה לא תספיק. האחריות המוסדית לפיתוח מסגרת משטרית ראויה לשימוש ציבורי בבינה מלאכותית מוטלת גם על המחוקק, על הממשלה, על הרגולטורים הענפיים ועל ספקי המערכות עצמם. המחוקק נדרש להכריע אם, ובאיזו מידה, נחוצה בישראל מסגרת חקיקתית אופקית או לכל הפחות עיגון מכורש של מנגנוני הגנה בסיסיים. הממשלה נדרשת להבטיח שהגישוה הסקטוריאלית שכבר הותוותה לא תישאר הצהרת מדיניות בלבד. הרגולטורים הענפיים נדרשים למלא בפועל את תפקיד המיפוי, הניתוח והפיקוח שהמדיניות הטילה עליהם, וספקי מערכות AI נדרשים להבין כי מכירה למגזר הציבורי אינה יכולה להתבסס על דרישה לאמון עיוור במערכת אטומה. ואולם, דווקא משום שכל אלה הם תהליכים איטיים, ולעיתים גם מפוצלים ומורכבים, בטווח הקצר מבקר המדינה עשוי להיות הגורם המעשי ביותר ליצירת שינוי. הוא אינו צריך להמתין להשלמת חקיקה ואינו חייב להמתין להסכמה מלאה בשאלה איזה מודל רגולטורי יאמץ המשפט הישראלי בעתיד.

די בכך שיבחן אם מוסדות המדינה, כאן ועכשיו, מפעילים מערכות המשפיעות על זכויות הפרט בלי שהוקמו סביבן אמצעי הבקרה המתבקשים. במובן זה, ביקורת המדינה יכולה לשמש מנגנון ביניים הכרחי בין מצב של אימוץ טכנולוגי מואץ לבין הסדרה משפטית שעדיין נמצאת בתנועה. הספרות הביקורתית על ממשל אלגוריתמי עומדת שוב ושוב

Eubanks, Automating Inequality 28, לעיל ה"ש 3, במיוחד הדיון בהליך הוגן, בשיעתוק אי-שוויון ו"בבית המחסה הדיגיטלי לעניים" (Digital poorhouse).

Pasquale, The Black Box Society 29, לעיל ה"ש 5, דיון בכוחן של מערכות "הקופסה השחורה", בחוסר השקיפות שלהן ובצורך בפיקוח משמעותי על מערכות דירוג, ניקוד וקבלת החלטות.



ביקורת על מערכות בינה מלאכותית מדריך למבקרי המגזר הציבורי

מוסדות הביקורת העליונים של ברזיל, פינלנד, גרמניה, הולנד, נורווגיה ובריטניה

תקציר

הבינה המלאכותית מסתמכות על למידת מכונה (machine learning), שבמסגרתה מודלים לומדים דפוסים מתוך נתונים כדי להשיג מטרות ספציפיות.

טכנולוגיות חדשות מלוות על פי רוב בסיכונים חדשים. בזמן שהחקיקה הייעודית בנושא עודנה מתגבשת, קיים צורך במנגנוני בקרה ובביצוע ביקורות בתחום זה. קהילת הבינה המלאכותית מתמקדת יותר ויותר בכללי אתיקה ובהשפעתה החברתית של הבינה המלאכותית. מאז פרסום מסמך זה לראשונה, הונהג רישוי למבקרים המתמחים בבינה מלאכותית².

אף שרשויות הגנת המידע פיתחו קווים מנחים ייעודיים ומסוגלות ליטול על עצמן תפקיד פיקוחי להגנת המידע האישי, רבים מן הסיכונים המזוהים עם יישומי בינה מלאכותית חורגים אל מעבר לסוגיות של מידע אישי. לדוגמה, מערכות בינה מלאכותית אטומות (opaque) עלולות להפוך פרקטיקות בלתי הוגנות לאוטומטיות ולחזק אותן, ובעקבות זאת לערער את האמון במוסדות ציבור. מוסדות ביקורת עליונים צריכים להיות ערוכים לבצע ביקורת של יישומי בינה מלאכותית באמצעות ביקורות ציות וביקורות ביצועים כאחד. כמה ממוסדות הביקורת העליונים החתומים על מזכר ההבנות ערכו חקרי מקרה (case studies) או ביקורות פיילוט³, במטרה לגבש מתודולוגיה גנרית לביקורת על יישומי בינה מלאכותית או לבחון את השימוש בבינה מלאכותית בשירות הציבורי.

המסמך הנוכחי מתווה את הסיכונים העיקריים הטמונים בשימוש בבינה מלאכותית בשירות הציבורי, ומציע שיטות לביצוע ביקורת על מערכות בינה מלאכותית. ההנחיות בו מבוססות על ניסיונם של מוסדות הביקורת העליונים שחיברו את המסמך (בברזיל, פינלנד, גרמניה, הולנד, נורווגיה ובריטניה), שכלל ביצוע ביקורות של למידת מכונה ופרייקטים בתחום התוכנה. זו איננה רשימת קריטריונים מחייבת, ויש להשתמש בה כמדריך לפיתוח או התאמה של מתודולוגיות הביקורת.

המסמך מתחלק לשני חלקים עיקריים. פרק 3 מציג קריטריונים אפשריים לביקורת, ובכללם תקנות לאומיות, תקנים בין-לאומיים וקווים מנחים הנמצאים בשימוש נרחב.

מסמך זה קובע הנחיות עבור ביקורת מערכות בינה מלאכותית (AI) המבוצעת על ידי מוסדות ביקורת עליונים (SAI). מטרתו לסייע למוסדות ביקורת עליונים ולמבקרים עצמם לבצע ביקורת של מערכות בינה מלאכותית המשמשות גופים במגזר הציבורי. הוא נועד למבקרים בעלי ידע מסוים בשיטות כמותיות, אך אינו מניח כי קיימת מומחיות בתחום של מערכות בינה מלאכותית או מודלים של למידת מכונה.

המסמך כולל קטלוג ביקורת, המורכב מקווים מנחים בנושאי ביקורת מוצעים מבוססי סיכונים, נוסף על בקורות מצופות. למסמך זה נלווה כלי עזר בפורמט Excel המסכם ומנחה את המבקרים בביצוע חלקים שונים של הביקורת.

1. תקציר מנהלים

בשנת 2017 חתמו מוסדות הביקורת העליונים של ברזיל, פינלנד, גרמניה, הולנד, נורווגיה ובריטניה על מזכר הבנות לשיתוף פעולה בנושא ניתוח נתונים. לאור ההכרה בטרנספורמציה הדיגיטלית המשנה את אופן פעולתם של גופי ציבור, הסכימו מוסדות הביקורת לשותף ידע ביניהם ולגבש גישות ביקורת חדשות. בשנת 2019 הם התחייבו להפיק במשותף מסמך הנחיות בנושא של בינה מלאכותית במגזר הציבורי. הגרסה הראשונה של מסמך זה פורסמה ב-2020 כסקירה טכנית, ועודכנה לראשונה ב-2023¹. המסמך עבר עדכון משמעותי לגרסת 2025 הנוכחית עקב שינויים משמעותיים בשימוש בבינה מלאכותית ובסיכונים המזוהים עימה, לצד התפשטות השימוש בבינה מלאכותית יוצרת, חקיקתן של תקנות אסדרה (רגולציה) חדשות וזמינות הנרחבת של מערכות בינה מלאכותית מסחריות.

בינה מלאכותית מצויה כיום בשימוש נרחב במגזר הציבורי, במטרה לשפר את השירותים ולהפחית עלויות. מערכות בינה מלאכותית יכולות לספק חיזויים (בינה מלאכותית לחיזוי), המלצות או החלטות, או לייצר תוכן כמו טקסט, תמונות או שמע (בינה מלאכותית יוצרת). מרבית מערכות

1 עדכון ראשון זה מתבסס על ניסיונם של המוסדות החתומים על מזכר ההבנות - ניסיון שנצבר דרך יישום המתודולוגיה שפותחה בביקורות שערכו אותם מוסדות.

2 ראו לדוגמה, הסמכת AAIA מטעם ISACA.

3 ראו לדוגמה את המסמכים הבאים: הבנת אלגוריתמים וכן ביקורת של 9 אלגוריתמים המשמשים את ממשלת הולנד, שפרסם בית המשפט לביקורת של הולנד; השימוש בבינה מלאכותית בממשלה המרכזית מטעם משרד הביקורת הלאומי של נורווגיה (גרסה מלאה בנורווגית בלבד); שיטות ניתוח נתונים ובינה מלאכותית בממשל הפדרלי הגרמני מאת בית הדין הפדרלי לביקורת חשבונות של גרמניה (בגרמנית בלבד); או שימוש בבינה מלאכותית בממשלה, שהפיק משרד הביקורת הלאומי הבריטי.

פרק 4 מציג קטלוג ביקורת⁴ הבנוי סביב תחומי מפתח: ניהול וממשל של הפרויקט, נתונים, פיתוח המערכת, הערכה לפי פריסה, פריסה וניהול שינויים, ומערכות בינה מלאכותית בסביבה תפעולית. עבור כל אחד מן התחומים אנו כוללים סקירה כללית, מצביעים על סיכונים ספציפיים לבינה מלאכותית ומציעים בקורות. אנו כוללים גם כלי עזר שיכול לשמש מבקרים בהכנות לביקורת.

עם האתגרים העיקריים שזיהו מוסדות הביקורת העליונים נמנים אלה:

- מפתחים עשויים להתמקד בביצועים טכניים, ולהתעלם מציות, שקיפות והוגנות.
- תקשורת גרועה בין בעלי המוצר למפתחים עלולה להוביל למערכות לא מועילות או יקרות.
- ארגונים רבים חסרים את המיומנויות לפיתוח בינה מלאכותית בתוך הארגון ומסתמכים על ספקים חיצוניים, מצב שיש בו כדי להגביר סיכוני ציות.
- שוררת אי-ודאות לגבי השימוש בנתונים אישיים בבינה מלאכותית, עם אחריותיות לא ברורה ומבנים ארגוניים מוגבלים.

המבקרים זקוקים להכשרה ספציפית בתחומי המומחיות הנמנים להלן כדי לבצע הערכות בעלות משמעות ביישומי בינה מלאכותית ולספק המלצות הולמות:

- ביקורות בסיסיות (baseline audits): כדי לסקור תיעוד באופן מועיל, המבקרים נדרשים להיות בעלי הבנה טובה של העקרונות הבסיסיים של מערכות בינה מלאכותית ואלגוריתמים של למידת מכונה, ובעלי ידע עדכני על ההתפתחויות הטכניות המהירות בתחום זה.
- ביקורות מעמיקות (in-depth audits): כדי לבצע ביקורות הכוללות בדיקות ביצועים משמעותיות, מבקרים זקוקים למיומנויות מעשיות בשפות תכנות נפוצות וביישום מודלים, וכן לידע וליכולת להשתמש בכלי תוכנה מתאימים.
- תשתית טכנולוגיית מידע (IT infrastructure): מאחר שמערכות בינה מלאכותית משתמשות לעיתים קרובות בפתרונות מבוססי ענן כדי למלא אחר דרישות המחשוב הגבוהות, על המבקרים להיות בעלי הבנה בסיסית בשירותי ענן ובאופן שבו הם תומכים בפעולות בינה מלאכותית.

מסמך זה מציע את המסקנות וההמלצות הבאות עבור

מוסדות ביקורת עליונים:

- על מוסדות ביקורת עליונים להיות ערוכים לבקר מערכות בינה מלאכותית כדי למלא את התפקידים המוקנים להם בחוק, ולהעריך אם השימוש במערכת המבוקרת מספק שירותים ציבוריים יעילים ומועילים המצייתים לחוק ולכללים.
- ביקורות בינה מלאכותית דורשות מן המבקרים לרכוש מיומנויות וידע מיוחדים. על מוסדות ביקורת עליונים להשקיע בכישורים של מבקריהם: מבקרים המבצעים ביקורות ביצועים, ציות או טכנולוגיית מידע צריכים להיות בעלי הבנה כללית טובה של סוגים שונים של מערכות בינה מלאכותית, מקרי השימוש שלהם, כללי אתיקה וסיכונים ספציפיים הנוגעים לבינה מלאכותית. לצורך בדיקות ביקורת טכניות ברמה שמעבר לניהול המערכת (ממשל), דרושים מבקרים בעלי מיומנויות בפיתוח בינה מלאכותית.
- קטלוג הביקורת וכלי העזר במסמך זה יושמו בהצלחה בכמה מקרי בוחן וניתן להתאימם לביקורות עתידיות.

המחברים מקווים כי מסמך הנחיות זה והכלים המלווים אותו יתמכו בקהילת הביקורת הבין-לאומית בהפקת ביקורות מועילות של בינה מלאכותית.

2. הקדמה

סוגי מערכות בינה מלאכותית ביישומי המגזר הציבורי

הבינה המלאכותית מתפתחת במהירות, והשימוש בה במגזר הציבורי הולך ומתרחב. גופים במגזר הציבורי מאמצים בינה מלאכותית כדי לשפר יעילות, לספק שירותים משופרים ולהפחית עלויות.

מערכות בינה מלאכותית הן טכנולוגיות מעשה ידי אדם שיועדו לבצע מטלות אשר באופן טיפוסי מצריכות אינטליגנציה (בינה). מערכות אלה לומדות מתוך נתונים ומתאימות את עצמן להשגת מטרות ספציפיות, כך שעל פי רוב הן משפרות את ביצועיהן בעצמן, ללא תכנות נוסף במפורש. בינה מלאכותית כוללת תחומים כמו למידת מכונה, עיבוד שפה טבעית, ראייה ממוחשבת, זיהוי קולי, תכנון ורובוטיקה. היא מאפשרת יכולות כמו תפיסה,

4. במושג קטלוג ביקורת כוונתנו למערכת קווים מנחים הכוללים הן את התוכן המוצע של נושאי ביקורת על בסיס סיכונים, הן את המתודולוגיה לביצוע בדיקות ביקורת הנוגעות לכך.

מערכות בינה מלאכותית לחיזוי (predictive AI) משתמשות בנתונים היסטוריים כדי לזהות דפוסים ולספק ניבויים לגבי מקרים חדשים. לדוגמה, הן מסוגלות להעריך את הסיכון להתפתחות מחלה על פי רשומות רפואיות

הנמקה, קבלת החלטות ופתרון בעיות. מרבית מערכות הבינה המלאכותית מסתמכות על למידת מכונה (machine learning), שבמסגרתה מודלים לומדים דפוסים מתוך נתונים כדי להשיג מטרות ספציפיות.

סיווג לפי טכנולוגיית בסיס, מורכבות ומקרי שימוש נפוצים

סוג מערכת בינה מלאכותית	טכנולוגיית בסיס	מורכבות	מקרי שימוש נפוצים במגזר הציבורי
מערכות מבוססות כללים	- מערכות מומחה - עצי החלטה - מנועי כללים עסקיים	נמוכה	- קביעת זכאות (כגון סינון לקבלת הטבות) - בדיקות ציות אוטומטיות (כגון אישור דוחות מס) - תמיכה מובנית בהחלטות
בינה מלאכותית לחיזוי (סיווג)	- רגרסיה לוגיסטית - יערות אקראיים - רשתות עצביות	בינונית	- גילוי מרמה (כגון סימון סיכון להונאת מס) - תעדוף מקרים (כגון מיון תביעות רווחה) - זיהוי חריגים (כגון סימון יעדים לביקורת מכס)
בינה מלאכותית לחיזוי (רגרסיה)	- רגרסיה לינארית - מודלים של סדרות עתיות - צירוף מודלים	בינונית	- חיזוי ביקוש (כגון ביקוש למיטות אשפוז) - הקצאת משאבים (כגון אומדני התחייבויות פנסיוניות עתידיות) - אומדן תקציב (כגון תחזיות צריכת אנרגייה)
ראייה ממוחשבת	- רשתות עצביות קונבולוציוניות - מודלים של גילוי אובייקטים - פילוח תמונות	גבוהה	- עיבוד מסמכים (כגון התאמה לתמונות הדרכון) - ביקורת תשתיות (כגון איתור פגמים בכביש) - מעקב וניטור (כגון ניתוח תמונות לוויין)
מודלי שפה גדולים (LLM)	- ארכיטקטורת טרנספורמר - מודלי יסוד מאומנים מראש - כוונון עדין / RAG	גבוהה מאוד	- טיפול בפניות האזרח (כגון צ'טבוטים בשירותים לציבור) - סיכום מסמכים (כגון ניתוח מסמכי מדיניות) - סיוע בניסוח (כגון ניסוח תכתובות)
בינה מלאכותית יוצרת (מודלית מודאלית)	- מודלי דיפוזיה - טרנספורמרים מולטי מודליים - רשתות יריבות יוצרות	גבוהה מאוד	- יצירת תוכן (כגון חומרי מידע לציבור) - סימולציה של הדרכה (כגון נתונים סינתטיים לאימון מודלים) - כלי נגישות (כגון תיאורי אודיו)

הערה: סיווג המורכבות משקף אתגרים ביכולת לבצע ביקורת במערכות אלה, לרבות היכולת לפרש את תוצאות המודלים, דרישות הנוגעות לנתונים והצורך במומחיות טכנית מיוחדת. מערכות בעלות רמת מורכבות גבוהה מצריכות על פי רוב בקורות ממשל ופיקוח קפדניות יותר.

לוח 2.1: סוגי מערכות בינה מלאכותית ביישומי המגזר הציבורי

או לסמן טעויות פוטנציאליות בדוחות מס. מערכות אלה משמשות באופן שכיח לסיווג (כמו מיון מקרים לקטגוריות) או גרסיה (כמו יבוי ערכים מספריים).

מערכות בינה מלאכותית יוצרת (generative AI) יודעות ליצור תוכן חדש (כמו טקסט, תמונות, וידאו או אודיו) על ידי למידה מתוך קובצי נתונים גדולים. הן מסוגלות לייצר תוצרים ריאליסטיים, על פי רוב עם הכוונה אנושית מעטה או ללא הכוונה כזו כלל.

מודלי יסוד (foundation models) או מודלי בינה מלאכותית למטרות כלליות (general-purpose AI models) מאומנים על קובצי נתונים גדולים ומגוונים מאוד. הם משתמשים בטכניקות למידה עמוקה (deep learning) וניתן להתאים אותם למגוון רחב של יישומים, מזיהוי תמונות ועד מחקר מדעי. **מודלי שפה גדולים (Large Language Models - LLMs)** הם אחד מהסוגים העיקריים של מודלי יסוד, המאומנים בעיקר על טקסטים. הם מסוגלים לבצע מטלות רבות כמו תרגום, סיכום ושיחה על ידי הפקת שפה הדומה לשפה אנושית.

השימוש בבינה מלאכותית בשירותים לציבור טומן בחובו הזדמנויות חדשות אך גם אתגרים, ובהם הצורך להגן על נתונים אישיים, הקפדה על כך שההחלטות תהיינה הוגנות וניתנות להסבר, והימנעות מהטיה ואפליה. מערכות בינה מלאכותית שאינן בנויות או מנוהלות כראוי עלולות להגביר את עומס העבודה, לגרום לעיכובים או לפגוע באמון הציבור.

כדי להתמודד עם סיכונים אלה, יש הכרח בבקורות פנימיות לצד ביקורות חיצוניות. גופים בין-לאומיים ולאומיים פיתחו עקרונות וקווים מנחים המסייעים לשימוש אחראי בבינה מלאכותית.⁵ עם התפתחות חקיקה ייעודית לבינה מלאכותית, מוקמים גופי רגולציה ופיקוח חדשים. לדוגמה, חוק הבינה המלאכותית של האיחוד האירופי (להלן - EU AI Act) מסדיר את הקמתן של רשויות ייעודיות לאכיפת ציות ולתמיכה בביקורת. רשויות אלה כוללות גופים מוסמכים⁶ ורשויות פיקוח שוק.⁷ לגופים אלה מוקנות זכויות גישה נרחבות לתיעוד הטכני או לרכיבי הבסיס של מערכות בינה מלאכותית, לרבות הנתונים וקוד המקור. מסגרות של פיקוח על בינה מלאכותית יכולות להוביל לקביעת תקנים

והסמכות מקצועיות, דבר שיקל על ביצוע ביקורות בתחום. מוסדות ביקורת עליונים (SAI) ממלאים תפקיד מרכזי בדרישת אחריות מגופי ציבור. בין היתר הם יכולים לדרוש שהשימוש בבינה מלאכותית יהיה בטיחותי ואתי. אומנם ביקורת בינה מלאכותית עלולה להיות כרוכה באתגרים, אך היא הולכת ונעשית חיונית יותר ככל שהטכנולוגיות הללו נפוצות יותר במגזר הציבורי. הסמכה למבקרי בינה מלאכותית מתמחים מורשים מפותחת בימים אלה.⁸

מסמך זה מציע הכוונה למוסדות ביקורת עליונים בכל הנוגע לביקורת על מערכות בינה מלאכותית המשמשות במגזר הציבורי. הוא מציע בהרחבה את הסיכונים המרכזיים לצד אמצעים להפחתתם, ומפרט אילו בקורות צריכים המבקרים לצפות למצוא בארגון המבוקר. ביקורות בינה מלאכותית יכולות לכלול אלמנטים של ביקורת ביצועים וכאלה של ביקורת ציית. זו איננה רשימת קריטריונים מחייבת, ויש להשתמש בה כמדריך לפיתוח או התאמה של מתודולוגיות הביקורת.

מערכות בינה מלאכותית כמעט לעולם אינן פועלות בנפרד; הן מוטבעות על פי רוב בתהליכי טכנולוגיית מידע (IT) רחבים יותר. על המבקרים להעריך תמיד את הסיכונים הקשורים לסביבת טכנולוגיית המידע הרחבה, והערכה זו עשויה להוביל לביקורת טכנולוגיית מידע מקיפה יותר. משימה זו עשויה להצריך מומחיות של אנשי מקצוע המתמחים בבינה מלאכותית או בטכנולוגיות מידע. מדריך זה מתמקד באופן ספציפי בביקורת של רכיב הבינה המלאכותית.

ההמלצות כאן שואבות ממחקרים שהתפרסמו ומניסיונם של מוסדות הביקורת העליונים שחיברו את המסמך, לרבות ביקורות של מערכות בינה מלאכותית ופרייקטים אחרים בתחום התוכנה. המסמך משקף את הניסיון בביקורות בינה מלאכותית באותם מוסדות ביקורת עליונים שחיברו את המסמך במועד הכתיבה⁹, וייתכן שהוא יעודכן עם הצטברותו של ניסיון ביקורת נוסף וקבלת תוצאות של מחקרים חדשים.

פרק 3 מציג קריטריונים אפשריים לביקורת, ובכללם תקנות לאומיות, תקנים בין-לאומיים וקווים מנחים הנמצאים

5 אנו מציינים אותם לצד מקורות נוספים לקריטריוני ביקורת בפרק 3 של מסמך זה.

6 EU AI Act, נספח VII; סעיף 3 (22): גוף מוסמך (notified body) הוא ארגון שהוסמך רשמית מכוח דיני האיחוד האירופי לבדוק אם מוצרים או מערכות מסוימים עומדים בתקנים הנדרשים.

7 EU AI Act, סעיף 74; סעיף 3 (26): רשות פיקוח שוק (market surveillance authority) היא הגוף הלאומי האחראי על ניטור מוצרים בשוק ואכיפת הציות לכללי האיחוד האירופי.

8 לדוגמה, ההסמכה של ISACA לתחום הבינה המלאכותית.

9 עדכון אחרון: דצמבר 2025.

עודן בתהליכי שינוי. מדינות שונות מגבשות תקנות אסדרה שמטרתן להנחות את הפיתוח של בינה מלאכותית ואת השימוש בה. פרק זה מציג סקירה כללית של הנוף האסדרתי במדינות שחיברו מסמך זה. הוא מציג גם תקנים וקווים מנחים בין-לאומיים שעשויים לשמש בידי המבקרים כדי להעריך אם משתמשים בבינה מלאכותית בצורה בטיחותית ומתוך שמירה על כללי האתיקה.

לא כל היבטי התכנון והפיתוח של מערכות בינה מלאכותית והשימוש בהן מאוסדרים באופן ברור. במקרים כאלה, המבקרים יכולים להיעזר בתקנות או בקווים מנחים קיימים מתחומים קרובים, כמו אבטחת מידע או הגנת מידע.

פרק זה מציג תחילה תקנות ויסודות משפטיים ייעודיים לתחום הבינה המלאכותית. בהמשך הוא מציין קווים מנחים אחרים, תקנים וגישות מומלצות (best practices), ומסביר בקצרה אילו דרישות מתחומים קרובים אחרים רלוונטיים בדרך כלל לבינה מלאכותית.

3.1 המסגרת המשפטית ליישומי בינה מלאכותית

מדינות שונות קובעות תקנות שמטרתן להנחות כיצד יש לפתח בינה מלאכותית וכיצד להשתמש בה. על המבקרים להבין את הנוף החקיקתי כדי להעריך אם ארגונים במגזר הציבורי מצייתים לחוק ומשתמשים בבינה מלאכותית באופן בטיחותי ואתי.

3.1.1 EU AI Act: האיחוד האירופי

EU AI Act¹⁰ הוא דבר החקיקה המקיף הראשון להסדרת תחום הבינה המלאכותית באיחוד האירופי. מטרתו לוודא שמערכות בינה מלאכותית בשוק האיחוד האירופי הן בטוחות לשימוש ומכבדות זכויות יסוד. הגישה שנוקט החוק מתמקדת בבטיחות המוצר, בדומה לכללים הנוגעים לטכנולוגיות אחרות הכרוכות בסיכון גבוה.

3.1.1.1 EU AI Act: תחולת ה-

החוק מגדיר בינה מלאכותית בצורה רחבה, אך אינו כולל מערכות מומחה פשוטות מבוססות כללים¹². הוא מציין במפורש כי מערכות בינה מלאכותית מסיקות כיצד להפיק פלט מהקלט שהן מקבלות¹³. ארגונים יכולים למלא תפקידים שונים על פי החוק; הם יכולים להיות "ספקים" (providers) אם הם מפתחים או מזמינים מערכות בינה מלאכותית, או

בשימוש נרחב. פרק 4 מציג קטלוג ביקורת¹⁰ הבנוי סביב תחומי מפתח: ניהול וממשל של הפרייקט, נתונים, פיתוח המערכת, הערכה לפני פריסה - כולל שיקולים אתיים כמו הוגנות והסברתיות (explainability - נקרא גם הסבריות), פריסה וניהול שינויים, ומערכות בינה מלאכותית בסביבה תפעולית. עבור כל אחד מן התחומים אנו כוללים סקירה כללית, מצביעים על סיכונים ספציפיים לבינה מלאכותית ומציעים בקורות.

למסמך זה מצורפים נספחים המכילים מידע נוסף:

3. נספח 1, רכיבים של ביקורת טכנולוגית מידע קלאסית בהקשר של בינה מלאכותית, מסכם את רכיבי הביקורת של טכנולוגית מידע הרלוונטיים לבינה מלאכותית.
 4. נספח 2, נתונים אישיים והתקנות הכלליות להגנת המידע (GDPR) בהקשר של בינה מלאכותית, כולל סיכונים הקשורים לנתונים אישיים ולהפרת ה-GDPR.
 5. נספח 3, מדדי שוויון והוגנות במערכות סיווג, מספק סקירה של מדדי השוויון וההוגנות הרלוונטיים ביותר ליישומי סיווג.
 6. נספח 4 מכיל מילון מונחים, בדגש על המונחים הטכניים המשמשים במסמך זה, לצד רשימת קיצורים וראשי תיבות המשמשים במדריך זה בנוסח המקורי שלו באנגלית.
 7. נספח 5 מכיל סיכום של תפקידים בגוף המבוקר העשויים להיות רלוונטיים לביקורת מערכת בינה מלאכותית.
- כלי עזר נפרד לביקורת עומד גם הוא לרשות הקוראים. כלי זה מציע מגוון של שאלות ביקורת הממופות לכל אחד משלבי מחזור החיים של בינה מלאכותית, ומסייע למבקרים לזהות ראיות ביקורת מתאימות.
- הכנת האמצעים החזותיים במדריך זה וכן עיבוד המדריך וכלי העזר לנוסח הסופי נעשו בחלקם בסיוע כלי בינה מלאכותית, עם בדיקות יסודיות על ידי המחברים.

3. מקורות של קריטריוני ביקורת

מאחר שבינה מלאכותית ולמידת מכונה מתפתחות ומשתנות במהירות, מסגרות החקיקה הקובעות את הכללים ליישומן

10 במושג קטלוג ביקורת כוונתנו למערכת קווים מנחים הכוללים הן את התוכן המוצע של נושאי ביקורת על בסיס סיכונים, הן את המתודולוגיה לביצוע בדיקות ביקורת הנוגעות לכך.

11 הדברים נאמרים מכורשות ב-Recital 12 של ה-EU AI Act.

12 הדברים נאמרים מכורשות ב-Recital 12 של ה-EU AI Act.

13 הנציבות האירופית פרסמה קווים מנחים לגבי ההגדרה של מערכת בינה מלאכותית.

השוואה בין מסגרות אסדרה
גישות מרכזיות לממשל בינה מלאכותית לפי אזור

האירופי - AI Act	ברזיל - PL 2338/2023	בריטניה - Pro-Innovation
גישת מרכזית סיווג מבוסס סיכונים המלווה בחובות תואמות קטגוריות סיכון • אסור (סיכון בלתי קביל) • סיכון גבוה (נדרש ציות לדרישות) • חובת שקיפות • סיכון מינימלי (אין דרישות) דרישות עיקריות • מערכות ניהול סיכונים • ממשל נתונים (נקרא גם משילות נתונים) • תיעוד טכני • פיקוח אנושי	גישת מרכזית מערכת שכבות סיכון המתואמות עם מסגרת האיחוד האירופי + הוראות חוק הגנת המידע על פי LGPD קטגוריות סיכון • סיכון מופרז • סיכון גבוה • סיכון בינוני • סיכון נמוך דרישות עיקריות • דרישות שקיפות • הסברתיות • פיקוח אנושי • הפחתת הטיות	גישת מרכזית מסגרת מבוססת עקרונות עם הנחיות ייעודיות למגזרים שונים עקרונות מרכזיים • בטיחות, אבטחה ועמידות • שקיפות והסברתיות • הוגנות • אחריותיות וממשל יישום • גופי האסדרה המגזריים מכניסים התאמות להנחיות • בקורות מידתיות • התמקדות בחדשנות • הערכות ספציפיות להקשר

הערה בנוגע לתחולת EU AI Act: ה-AI Act אינו חל על מערכות בינה מלאכותית המשמשות למטרות הגנתיות וצבאיות, למטרות ביטחון לאומי או למטרות מחקר ופיתוח.

סיכון גבוה יותר = נדרש תיעוד מקיף יותר וראיות לציות. על המבקרים להעריך אילו מסגרות אסדרה חלות, על פי תחום השיפוט וההקשר של פריסת המערכת.

לוח 3.1: השוואה בין מסגרות האסדרה המתוארות במדריך זה.

בינה מלאכותית לארבע קטגוריות¹⁵. איור 3.1 מציג את הקטגוריות השונות של מערכות בינה מלאכותית המוגדרות ב-EU AI Act.

- **בינה מלאכותית אסורה בשימוש:** שימושים מסוימים אסורים בחוק מלכתחילה, כמו דירוג חברתי, פרופילינג לצורכי חיזוי, וזיהוי רגשות במקומות עבודה או בבתי ספר¹⁶.
- **בינה מלאכותית בסיכון גבוה:** עם אלה נמנות, לדוגמה, מערכות בינה מלאכותית המשמשות

"מפעילים" (deployers) אם הם משתמשים במערכות בינה מלאכותית במסגרת פעילותם.

החוק חל על מרבית מערכות הבינה המלאכותית אשר מוצעות או נמצאות בשימוש באיחוד האירופי, אך יש חריגים מסוימים¹⁴. לדוגמה, הוא אינו מכסה מערכות המשמשות באופן בלעדי לצורכי ביטחון לאומי, הגנה או מחקר מדעי.

3.1.1.2 גישת מבוססת סיכונים

החוק מחיל מסגרת מבוססת סיכונים. הוא מסווג מערכות

14 EU AI Act, סעיף 2.

15 הנציבות האירופית מספקת כלי לבדיקת ציות. מבקרים יכולים להשתמש בכלי זה כדי להעריך אילו הוראות של ה-AI Act חלות על מערכת בינה מלאכותית נתונה.

16 הנציבות האירופית פרסמה קווים מנחים לגבי בינה מלאכותית אסורה בשימוש.

- יש להבטיח אוריינות בינה מלאכותית (AI literacy). ספקים ומפעילים חייבים לפעול כדי לוודא שהמשתמשים בעלי כשירות מספקת להשתמש במערכות בינה מלאכותית.

ה-EU AI Act מעודד קביעת קודי התנהלות מרצון המחילים את האסדרה של מערכות בסיכון גבוה על מערכות אחרות.

3.1.1.4 מודלים של בינה מלאכותית למטרות כלליות

החוק קובע כללים מיוחדים למודלים של בינה מלאכותית למטרות כלליות (general-purpose AI models).

מודלים עוצמתיים במיוחד מסוג זה מסווגים כמערכות הכרוכות בסיכון מערכתי (systemic risk). סיווג זה מבוסס על יכולת ההשפעה של המודל. מודל ייחשב ככרוך בסיכון מערכתי אם הוא אומן על ידי שימוש ביותר מ-1025 פעולות חישוב בנקודה צפה (FLOPs).

חשוב להבחין בין מודל בינה מלאכותית למטרות כלליות (כמו GPT-4) לבין מערכת שבנויה על גבי מודל כזה (כמו ChatGPT). האסדרה לגבי בינה מלאכותית למטרות כלליות חלה בעיקר על המודל עצמו¹⁹. מודלים של בינה מלאכותית למטרות כלליות אינם מסווגים כמערכות בינה מלאכותית החוסות תחת החוק, כך שהם אינם כפופים לכללים הנוגעים למערכות אסורות בשימוש או בסיכון גבוה. עם זאת, אם מודל למטרות כלליות משמש כחלק ממערכת בינה מלאכותית בסיכון גבוה, הספק והמשתמש של אותה מערכת חייבים לוודא שהמערכת עונה על כל הדרישות הרלוונטיות. ספקים של מודלי בינה מלאכותית למטרות כלליות מחויבים לספק מידע ותיעוד על המודל²⁰. הדבר נועד לסייע למשתמשים להבין את יכולות המודל ומגבלותיו, ולסייע לארגונים לציית לחוק²¹.

3.1.1.5 החלת ה-EU AI Act

ה-EU AI Act נכנס לתוקף באוגוסט 2024. הוראות החוק עתיד להיות מיושמות בהדרגה על פני כמה שנים. עבור מערכות בינה מלאכותית שכבר פועלות, הדרישות הרלוונטיות עשויות לחול באופן חלקי בלבד או במועד מאוחר יותר.

במערכות אכיפת החוק, בתשתיות קריטיות (כולל מכשירים רפואיים) ובגישה לשירותים חיוניים. רבות מהוראות החוק מתמקדות במערכות אלה.

- **בינה מלאכותית עם חובת שקיפות:** מערכות שבאות במגע ישיר עם אנשים חייבות להיות מסומנות בבירור כבינה מלאכותית, ללא קשר לדרגת הסיכון.
- **מערכות בינה מלאכותית אחרות:** רוב מערכות הבינה המלאכותית האחרות אינן כפופות לכללים ספציפיים, אך עדיין חייבות לעמוד בחוקים הכלליים של האיחוד האירופי.

3.1.1.3 דרישות

במקרה של מערכות בינה מלאכותית בסיכון גבוה, אלה הדרישות מן הספקים¹⁷:

- הערכה ותיעוד של הסיכונים.
- ניהול רשומות (records) ויומני רישום (log) מפורטים של המערכת.
- פיקוח אנושי.
- עמידה בסטנדרטים של איכות נתונים, עמידות, אבטחת סייבר ודיוק.
- רישום מערכות בסיכון גבוה במסד נתונים של כלל האיחוד האירופי.

הנציבות האירופית, עם המדינות החברות, עתידה להקים מסד נתונים למערכות בסיכון גבוה. הוא יכלול תיאורים של יישום הבינה המלאכותית, של הנתונים ושל הפונקציונליות של המערכת, וכן מדריכים למשתמשים. תחול חובה לרשום במסד הנתונים גם מערכות שאינן מסווגות כמערכות בסיכון גבוה אך ורק בשל פטור.

עבור כל המערכות, בכל סיווגי הסיכון:

- אם מערכות בינה מלאכותית מקיימות מגע ישיר עם פרטים, חובה לסמן את המערכות ואת הפלטים שלהן כ"בינה מלאכותית" או "נוצר על ידי בינה מלאכותית"¹⁸.

17 ראו פרטים נוספים בפרק 3 ל-EU AI Act. הנציבות האירופית עתידה לפרסם קווים מנחים לגבי מערכות בינה מלאכותית בסיכון גבוה (בנובמבר 2025).

18 הנציבות האירופית עתידה לפרסם קווים מנחים לגבי חובות שקיפות אלה (בדצמבר 2025).

19 לקבלת הקשר נוסף, ראו: S. Wachter, Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the European Union, the United States, and Beyond", *Yale Journal of Law & Technology*, 26, 3 (2024).

20 הנציבות האירופית פרסמה כללי נוהל (Code of practice) לעבודה על מודלים של בינה מלאכותית למטרות כלליות.

21 לפי EU AI Act, סעיף 53, 2, חובות אלה אינן חלות על ספקים של מודלי בינה מלאכותית למטרות כלליות המתפרסמים תחת רישיון קוד פתוח או רישיון חנים ואינם יוצרים סיכונים מערכתיים.

ה-AI Act הוחל על ידי מדינות EFTA של האזור הכלכלי האירופי (נורווגיה, ליכטנשטיין ואיסלנד) באיחור מסוים.

3.1.2 החקיקה וההנחיות בבריטניה

בריטניה מסתמכת על חוקים מגזריים ועל גופי אסדרה כדי לנהל בקרה על פיתוח בינה מלאכותית ושימוש בה במגזר הציבורי, וזאת במקום חקיקה רוחבית ייעודית לבינה מלאכותית. בשנת 2023 פרסמה הממשלה סקירה טכנית בשם "גישה מקדמת חדשנות לאסדרת בינה מלאכותית" (A pro-innovation approach to AI regulation)²² המתווה עקרונות לגופי האסדרה. מסמך זה מתאר "חמישה עקרונות" לפיתוח ואסדרה של בינה מלאכותית. ואלה הם: בטיחות, אבטחה ועמידות (robustness); שקיפות נאותה והסבריות (explainability); הוגנות (fairness); אחריות (accountability) וממשל (governance); וכן חופש כניסה ויציאה (contestability) ופיצוי (redress). העקרונות מתארים מה חייבות מערכות בינה מלאכותית המשמשות במגזר הציבורי להשיג, וכן את הדרכים להשגת תוצאות אלה בפועל. מבקרים יכולים להשתמש בהנחיות הללו כדי להבין כיצד עשויות גישות מומלצות להיראות בשטח לאחר פריסת מערכת בינה מלאכותית.

רשויות קיימות מופקדות על יישום עקרונות אלה במגזרים שלהן, על פיתוח הנחיות אסדרה ועל הפיקוח על פריסת הטכנולוגיה. לדוגמה, נציבות המידע הבריטית (Information Commissioner's Office - ICO) עדכנה את חוק הנתונים (Data [Use and Access] Act) כך שיחול על נסיבות שבהן ארגון עשוי למכך קבלת החלטות (להפוך אותן לאוטומטיות) על בסיס עיבוד אוטומטי של מידע אישי²³.

ממשלת בריטניה מספקת עקרונות והנחיות למגזר הציבורי. לדוגמה, "המדריך לבינה מלאכותית במגזר הציבורי הבריטי" (AI Playbook for the UK Government) מספק הנחיות לשימוש בבינה מלאכותית באופן בטיחותי, מועיל ומאובטח על ידי עובדי המינהל הציבורי וארגונים ממשלתיים. הוא מתווה עשרה עקרונות לשימוש בבינה מלאכותית

בממשלה ובמגזר הציבורי. מבקרים יכולים להשתמש בהנחיות אלה כדי לדעת מה הן הגישות המומלצות לרכש ופריסה של מערכות בינה מלאכותית²⁴. כמו כן, התקן לתיעוד שקיפות אלגוריתמית (Algorithmic Transparency Recording Standard - ATRS) הוא כעת תקן מחייב עבור משרדי הממשלה²⁵. ה-ATRS קובע שיטה אחידה עבור ארגוני המגזר הציבורי לפרסום מידע על אופן השימוש שלהם בכלים אלגוריתמיים ועל הסיבות לשימוש זה.

הממשלה הבריטית גם הרחיבה הנחיות קיימות בנושא וכללה בהן המלצות לגבי מערכות בינה מלאכותית. לדוגמה:

3. *The AQUA Book* קובע הנחיות לביצוע ניתוח מהימן התואם לצרכים העסקיים. הוא מתווה גורמים לקביעת רמת הבקרה והאימות הנדרשת, כולל הנחיות ספציפיות להבטחת איכות של מערכות בינה מלאכותית²⁶. עם גורמים אלה נמנים קריטריונים עסקיים; חדשנות הגישה; רמת הדיוק הנדרשת בתוצרים; כמות המשאבים הזמינים לביצוע הבטחת האיכות; והשפעות על הציבור. המסמך כולל גם שיקולים בעיצוב מודלים של בינה מלאכותית, לרבות הערכת סיכונים, הערכת השפעות, ביקורת הטיות וביקורת ציות.

4. ה-*Magenta Book* קובע הנחיות להערכת השפעתן של התערבויות בינה מלאכותית²⁷.

5. המרכז הלאומי לאבטחת סייבר (National Cyber Security Centre) קובע עקרונות לאבטחת מערכות למידת מכונה²⁸.

6. קיים גם מסמך הנחיות לרכש בינה מלאכותית *Guidelines for AI Procurement*. המתאר גישות מומלצות לרכישת טכנולוגיות בינה מלאכותית בגופים ממשלתיים²⁹.

נוסף על כך קיימות הנחיות לאישור ולהבטחת איכות של בינה מלאכותית:

7. המסמך *Introduction to AI Assurance* של ממשלת בריטניה מביא סקירה כללית של מושגים בהבטחת

22 למידע נוסף ראו: A pro-innovation approach to AI regulation - GOV.UK.

23 למידע נוסף ראו: [Data protection | ICO](#).

24 המדריך נגיש כאן: [AI Playbook for the UK Government - GOV.UK](#).

25 למידע נוסף ראו: [Algorithmic Transparency Recording Standard Hub - GOV.UK](#).

26 למידע נוסף ראו: [Design - The AQUA Book 7](#).

27 למידע נוסף, ראו: [Guidance on the Impact Evaluation of AI Interventions \(HTML\) - GOV.UK](#).

28 למידע נוסף, ראו: [Principles for security of Machine learning ML - NCSC.GOV.UK](#).

29 למידע נוסף, ראו: [Guidelines for AI procurement - GOV.UK](#).

מדריך ה-TCU קובע כללים ברורים:

- **פתרונות בינה מלאכותית מאושרים:** רק כלי בינה מלאכותית מסוימים מאושרים לשימוש עם מידע סודי. כלים שאינם מאושרים על ידי TCU, כמו Gemini ו-ChatGPT, מותרים לשימוש רק עם נתונים ציבוריים ובתנאים מחמירים.
 - **אחריותיות המשתמשים:** עובדי הארגון ימשיכו לשאת באחריות מלאה לכל תוכן שמופק בעזרת בינה מלאכותית, גם אם הכלי הפיק טעויות.
 - **פיקוח אנושי:** כל ההחלטות האוטומטיות מחייבות סקירה אנושית, במיוחד כשמדובר במידע אסטרטגי או כזה המיועד לציבור.
 - **אימות תוכן (ולידציה):** תוכן שנוצר באמצעות בינה מלאכותית חייב להיבדק בהיבטים של דיוק, הוגנות והלימות. עובדי הארגון אינם רשאים להשתמש בבינה מלאכותית אם אין להם אפשרות לאמת את הפלט.
 - **אבטחה וציות:** על השימוש בבינה מלאכותית להיות תואם למדיניות אבטחת המידע של TCU. אי-ציות עלול להוביל להליך משמעתי או הפרת חוזה.
 - **קניין רוחני:** עובדי הארגון אינם רשאים להשתמש בתוכן שנוצר באמצעות בינה מלאכותית אם הוא עלול להפר זכויות יוצרים, ועליהם לסמן חומר שנוצר באמצעות בינה מלאכותית במקרה המתאים.
 - **פיתוח פתרונות חדשים:** כלי בינה מלאכותית יוצרת (generative AI) חדשים חייבים באישור יחידת טכנולוגיית המידע (IT) וועדת ההנהלה. רק יחידת ניהול ה-IT רשאית לפתח יישומי בינה מלאכותית המיועדים לקהלים חיצוניים.
- המדריך שם דגש גם על סיכוני ההטיה הטמונים בבינה מלאכותית. הוא מבחין בין הטיית מודל (מנתוני האימון) לבין הטיית אוטומציה (הנטייה האנושית לתת אמון בהצעות אוטומטיות). המדריך מדגיש כי כל שימוש בבינה מלאכותית חייב להתאים לקוד ההתנהלות של TCU ולמדיניות נגד אפליה.

איכות למערכות בינה מלאכותית. הוא אינו כולל פרטים טכניים להבטחת איכות של בינה מלאכותית, אך זהו מקור שימושי למידע רקע³⁰.

8. המסמך *AI in Audit* שפורסם מטעם המועצה לדיווח פיננסי (Financial Reporting Council) מציג חקר מקרים לבינה מלאכותית בביקורת לצד הנחיות לתיעוד כלי בינה מלאכותית³¹.

9. הנחיות הממשלה הבריטית לבקרת הוצאות דיגיטל וטכנולוגיה כוללות הנחיות לגבי רכש בינה מלאכותית במסגרת *Risk and Importance*³².

3.1.3 חקיקה והנחיות בברזיל

ברזיל מפתחת מסגרת חוקית לאסדרת בינה מלאכותית. הצעת חוק מס' 2338 משנת 2023³³ שהוגשה לסנאט קובעת כללים לפיתוח בינה מלאכותית ושימוש בה ברחבי המדינה. הצעת החוק מכסה טווח רחב של נושאים, כולל פיתוח כלכלי, מדעים וטכנולוגיה, הדין האזרחי וזכויות הפרט.

הסנאט אישר את הצעת החוק והיא נבחנת כעת על ידי בית הנבחרים. מחוקקים הציעו תיקונים אחדים. אלה כללו דרישות לסימון תוכן שנוצר באמצעות בינה מלאכותית, קידום מחקר והכשרה, הגדלת הענישה על הכפשה באמצעות בינה מלאכותית והגנה על ילדים ומתבגרים. תיקונים נוספים מתייחסים לפיתוח בינה מלאכותית ושימוש בה, שינויים בחוק הגנת המידע הכללי (LGPD) של ברזיל, השתתפות הציבור באסדרה והחובות החוקיות החלות על מפתחי בינה מלאכותית. הצעת החוק גם מכסה נושאים של זכויות יוצרים, קניין רוחני, זכויות עובדים וחופש הביטוי. תחולת הצעת החוק היא רחבת היקף. היא מטפלת בנושאים של אחריות, בטיחות, אמון, כיבוד זכויות אדם, דמוקרטיה, הגנת הסביבה, פיתוח בר קיימה, סיווג סיכונים, ממשל תאגידי, הערכת השפעות, חבות אזרחית ועונשים.

מוסד הביקורת העליון של ברזיל (Federal Court of Accounts - TCU) פרסם גם הנחיות מעשיות לשימוש בבינה מלאכותית. ביולי 2024 פרסם TCU מדריך לשימוש בבינה מלאכותית יוצרת. המדריך מיועד לכל העובדים, הקבלנים והמתמחים ב-TCU. מטרתו לתמוך בחדשנות ופרודוקטיביות ובה בעת להבטיח ביטחון נתונים, פרטיות ומהימנות³⁴.

30 למידע נוסף, ראו: [Introduction to AI assurance. - GOV.UK](#)

31 למידע נוסף, ראו: [AI in Audit](#)

32 למידע נוסף, ראו: [Risk and Importance Framework - GOV.UK](#)

33 [PL 2338/2023](#)

34 למידע נוסף, ראו: [Guia de uso de Inteligencia Artificial no Tribunal de Contas da Uniao \(TCU\)](#)

3.2 תקנים, קווים מנחים וגישות מומלצות

ארגונים שונים כבר הגדירו תקנים וקווים מנחים עבור היבטים שונים בפיתוח בינה מלאכותית.

הארגון לשיתוף פעולה ופיתוח כלכלי (OECD) פרסם עקרונות לבינה מלאכותית (AI Principles) ב-2019 ועדכן אותם ב-2024.³⁵ עקרונות אלה מצביעים על חמישה ערכי ליבה: שקיפות, הסבריות, עמידות, בטיחות ואבטחה. הארגון גם מספק המלצות, למשל השקעה במחקר ופיתוח בתחום הבינה המלאכותית.

המכון הלאומי לתקנים וטכנולוגיה של ארה"ב (National Institute of Standards and Technology - NIST AI Risk Management Framework) ב-2023.³⁶ מסגרת זו מסייעת לארגונים לזהות, להעריך ולנהל סיכונים המזוהים עם בינה מלאכותית. היא מכסה נושאים כמו ביסוס תרבות של מודעות לסיכונים, ניתוח סיכונים ופיתוח בקורות נאותות.

קבוצת מומחים לבינה מלאכותית באיחוד האירופי (EU High-Level Expert Group on Artificial Intelligence) פרסמה קווים מנחים אתיים לבינה מלאכותית מהימנה.³⁷ קווים מנחים אלה דורשים ממערכות בינה מלאכותית לעמוד בדרישות של חוקיות, אתיקה ועמידות. הם כוללים שבע דרישות לבינה מלאכותית מהימנה ורשימת תיג שתכליתה לסייע לארגונים להעריך את מידת עמידתם בדרישות.

מכון פראונהופר לניתוח בינה ומערכות מידע (The Fraunhofer Institute for Intelligent Analysis and Information Systems - IAIS) תרגם דרישות אלה לקטלוג פרקטי של הערכות, המקל על ארגונים ליישם את הקווים המנחים בשטח.³⁸

בשנת 2024 אימצה המועצה האירופית את אמנת המסגרת לבינה מלאכותית וזכויות אדם, דמוקרטיה ושלטון החוק (Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law).³⁹

האמנה דורשת מן המדינות החתומות עליה להגן על זכויות אדם, על תהליכים דמוקרטיים ועל שלטון החוק לאורך כלל מחזור החיים של מערכות בינה מלאכותית.

3.3 מקורות של קריטריוני ביקורת מתחומים משיקים

לא כל היבטי הבינה המלאכותית מוסדרים באסדרה ייעודית. במקרים של העדר אסדרה, המבקרים נדרשים לבחון תקנים ודרישות מקובלים מתחומים קרובים. אלה יכולים לספק יסודות חזקים להערכת שימוש בטוח ואחראי בבינה מלאכותית.

3.3.1 אבטחת מידע וממשל IT

מערכות בינה מלאכותית מעבדות לעיתים קרובות נתונים רגישים או תומכות בפעילות הקריטית של מוסד. לכן סטנדרטים של אבטחת מידע רלוונטיים במידה רבה. המבקרים יכולים לבדוק את הציות לתקני אבטחת מידע רלוונטיים - לאומיים או בין-לאומיים. לדוגמה, משרד אבטחת המידע הפדרלי הגרמני (BSI) פיתח גישה הנקראת "IT-Grundschutz". היא כוללת כמה פרסומים עיקריים⁴⁰:

תקציר (קומפנדיום) של "IT-Grundschutz", המתאר איומים פוטנציאליים ודרישות אבטחה.

תקנים של ה-BSI, שבהם מומלצים תהליכים, שיטות ומדדים של אבטחת מידע.

בדומה לכך, מסגרות כמו COBIT של ISACA יכולים להנחות ממשל IT ולסייע להבטיח את קיומן של בקורות תקפות.⁴¹

3.3.2 שימוש בשירותי ענן חיצוניים

מערכות בינה מלאכותית רבות, במיוחד אלה המשתמשות במודלי שפה גדולים (LLM), מסתמכות על שירותי ענן חיצוניים. שירותים אלה אומנם מציעים גמישות וסקלביליות (יכולת הרחבה), אבל הם גם יכולים להכניס סיכונים חדשים למערכת.

35 לאחר פרסום מסמך עקרונות הבינה המלאכותית (AI principles) המקורי, פורסם גם ארגז כלים (Toolkit) של מדעי הנתונים, שנועד לסייע ביישום העקרונות לתחום הבינה המלאכותית.

36 למידע נוסף, ראו: [AI Risk Management Framework](#).

37 High-Level Expert Group on AI (2019): [Ethics Guidelines for Trustworthy Artificial Intelligence](#).

38 למידע נוסף, ראו: [AI assessment catalogue](#).

39 למידע נוסף, ראו: [Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law](#).

40 למידע נוסף, ראו: [IT-Grundschutz](#).

41 למידע נוסף, ראו: ISACA, *COBIT 2019 Framework: Governance and management objective* (2018).

באמצעותן. לדוגמה, סעיף 41(2c) למגילת זכויות היסוד של מדינות האיחוד האירופי (Fundamental Rights of the European Union) קובע כי חובתם של גופי ממשל לנמק את החלטותיהם. ברמה הלאומית, הזכות לשקיפות והסברתיות מעוגנת לא פעם בחוקי חופש המידע.

בדומה לחוקי הגנת המידע, חוקים נגד אפליה מגינים על זכויות יסוד, ועל כן חלים גם על יישומי בינה מלאכותית. עם זאת חוקים אלה מנוסחים פעמים רבות בדרך שמגינה על הפרט מפני אפליה במקרים פרטניים או במצבים ספציפיים. עשויה להידרש פרשנות רחבה כדי שיחולו על אפליה אלגוריתמית המעמידה קבוצות אנשים בעמדת נחיתות סטטיסטית.

3.3.5 אתיקה בבינה מלאכותית

דרישות אתיקה רבות הנוגעות לבינה מלאכותית מושרשות בזכויות יסוד כמו פרטיות, הוגנות וכבוד האדם. קווים מנחים בין-לאומיים יכולים לסייע למבקרים לגבש קריטריונים לבינה מלאכותית אתית. קווים מנחים אלה מבוססים על פי רוב על אמנות לזכויות אדם ועל חוקות של מדינות שונות. לדוגמה, קבוצת המומחים לבינה מלאכותית באיחוד האירופי ביססה חלקית את הקווים המנחים האתיים לבינה מלאכותית מהימנה על אמנות האיחוד האירופי, על מגילת זכויות היסוד של האיחוד האירופי ועל הדין הבין-לאומי לזכויות אדם⁵⁰. גישה זו יכולה לשמש מודל לפיתוח קריטריוני ביקורת לבינה מלאכותית אתית, המעוגנים בזכויות אדם בסיסיות שניתן למצוא בחוקות רבות ובמסמכי התחייבות בין-לאומיים⁵¹.

פרסום ה-BSI "שימוש מאובטח בשירותי ענן" (*Secure Use of Cloud Services*)⁴² מתאר כיצד אפשר להשתמש בשירותי ענן בבטחה. כדי לעיין בהערכות טכניות של שירותי ענן המבקרים יכולים להיעזר במסגרות שפרסם ה-BSI, כגון "קטלוג בקורות ציות במחשוב ענן" (*Cloud Computing Compliance Controls Catalogue - C5*)⁴³ ו"קטלוג קריטריונים לשירותי ענן לבינה מלאכותית" (*AI Cloud Services Criteria Catalogue - AIC4*)⁴⁴. קיים גם המסמך "מטריצת בקורות ענן" (*Cloud Controls Matrix - CCM*) שפרסמה הברית לאבטחת ענן (*Cloud Security Alliance - CSA*)⁴⁵.

3.3.3 הגנת המידע

מערכות בינה מלאכותית משתמשות על פי רוב במאגרי נתונים גדולים העשויים לכלול נתונים אישיים. דיני הגנת המידע, כמו התקנה הכללית להגנת מידע (GDPR) של האיחוד האירופי, חלים הן על הפיתוח של בינה מלאכותית, הן על השימוש בה. כמה רשויות לאומיות להגנת המידע פרסמו קווים מנחים להגנת מידע ביישומי בינה מלאכותית; למשל ה-ICO בבריטניה⁴⁶, רשות הגנת המידע הנורווגית (Datatilsynet)⁴⁷, חוק הגנת המידע הכללי של ברזיל⁴⁸ וכן ועידת הגנת המידע הגרמנית (DSK)⁴⁹. נספח 2 מסכם כמה מן האתגרים המרכזיים הקשורים לציות ל-GDPR עבור מערכות בינה מלאכותית.

3.3.4 שקיפות ואי-אפליה

ארגוני המגזר הציבורי מחויבים לנהוג בשקיפות לגבי תהליכי קבלת החלטות. זכותם של האזרחים להבין כיצד מתקבלות החלטות המשפיעות עליהם, לרבות החלטות שנעזרות במערכות בינה מלאכותית ואף מתקבלות

42 למידע נוסף, ראו: [Secure Use of Cloud Services](#).

43 למידע נוסף, ראו: [Cloud Computing Compliance Controls Catalogue \(5\)](#).

44 למידע נוסף, ראו: [AI Cloud Services Criteria Catalogue \(AIC4\)](#).

45 למידע נוסף, ראו: [Cloud Controls Matrix](#).

46 הנחיות מטעם משרד נציב המידע: [UK GDPR guidance and resources](#) וכן [Guidance on AI and data protection](#).

47 למידע נוסף, ראו: [Artificial intelligence and privacy](#).

48 למידע נוסף, ראו: [LEI N° 13.709](#).

49 למידע נוסף, ראו: [Guidance on Recommended Technical and Organisational Measures for the Development and Operation of AI Systems](#) וכן [Guidance on Artificial Intelligence and Data Protection](#) (בגרמנית בלבד).

50 למידע נוסף, ראו: [Ethics Guidelines for Trustworthy AI](#).

51 משרד הביקורת הלאומי של נורווגיה ביסס קריטריוני ביקורת לפרטיות ואוטונומיה של האדם על האמנה האירופית לזכויות אדם, בשילוב עם החוקה הנורווגית. הוא פרסם את הקריטריונים בדוח ביקורת על השימוש בבינה מלאכותית בממשלה המרכזית (בנורווגיה בלבד); דוח מסכם ללא קריטריוני הביקורת זמין באנגלית).

3.3.6 קניין רוחני והשפעה סביבתית

השימוש בבינה מלאכותית, ובמיוחד השימוש במודלי שפה גדולים (LLM), מעורר שאלות לגבי קניין רוחני. אימון נתונים עשוי לכלול חומרים הכפופים לזכויות יוצרים. ה-EU AI Act והדירקטיבה לזכויות יוצרים של האיחוד האירופי (EU Copyright Directive)⁵² הם דברי חקיקה המסדירים דיני זכויות יוצרים ביחס לבינה מלאכותית במדינות האיחוד האירופי. בריטניה דורשת מארגונים לציית לדין הבין-לאומי לזכויות יוצרים ושוקלת לעדכן את מדיניותה כדי לספק בהירות לגבי תחולתו על בינה מלאכותית.⁵³

מערכות בינה מלאכותית יכולות גם להשפיע משמעותית על הסביבה, בפרט אלה שמצריכות כוח מחשוב רב. יש תקנות הדורשות מארגונים לדווח על טביעת הרגל הסביבתית של מערכות הבינה המלאכותית שלהם. לדוגמה, ה-EU AI Act מעודד שימוש בבינה מלאכותית בדרך שממזערת פגיעה בסביבה⁵⁴, וקובע כי במקרה של מערכות בינה מלאכותית מסוימות, חלה דרישה לספק מידע על ההשפעה הסביבתית מבחינת צריכת האנרגיה. בבריטניה יש ארגונים במגזר הציבורי הנדרשים לתת גילוי נאות למידע מסוים על קיימות, וצפוי כי הנחיה זו תורחב כדי לכלול את ההשפעות של בינה מלאכותית. הערכת ההשפעה הסביבתית של בינה מלאכותית היא תחום חדש הראוי לתשומת לב בהקשר של ביקורת. אף שהשיטות למדידת השפעה זו עודן בתהליכי פיתוח, כלים מסוימים כבר הותוו על ידי ה-OECD.⁵⁵

4. קטלוג ביקורת

קטלוג הביקורת המתואר בפרק זה כולל קווים מנחים לביקורת מערכות בינה מלאכותית, הכוללים הן את התוכן המוצע של נושאי ביקורת על בסיס סיכונים, הן את הבקורות או האמצעים להפחתת הסיכונים שעל המבקרים לצפות לראות בגוף המבוקר.

מערכות בינה מלאכותית מוטמעות על פי רוב בתשתיות ID רחבות יותר. אף שיישומי בינה מלאכותית כרוכים באתגרים שונים מאלה המזוהים עם מערכות תוכנה או ID מסורתיות, עדיין ניתן לארגן את הביקורת על יישומי בינה מלאכותית ומודלי למידת מכונה בעזרת מסגרות עבודה

המוכרות למבקרי IT - כמו התהליך התקני חוצה התעשיות לכריית מידע (Cross-Industry Standard Process for Data Mining), הידוע בשמו המקוצר CRISP-DM.⁵⁶ המסמך הנוכחי מתמקד בסוגיות הנוגעות במפורש ליישומי בינה מלאכותית, ובפרט בשימוש בטיחות, אחראי ואטי בבינה מלאכותית. השיקולים העיקריים כוללים שקיפות והסברתיות של החלטות מערכות הבינה המלאכותית, שוויון והוגנות של החלטות אלה ואבטחת המערכות. על היבטים אלה להיות משולבים בתהליך הפיתוח לכל אורכו, בהתאם לעקרונות של הוגנות מובנית בתכנון ("by design"), שקיפות מובנית בתכנון ופרטיות מובנית בתכנון. לפיכך, אין ייעודו של קטלוג ביקורת זה לשמש כרצף לינארי של צעדים יחידים, אלא שהוא מייצג לולאה רציפה של שלבים המקיימים אינטראקציה ביניהם.

4.0 עומק הביקורת ומבנה הביקורת

ביקורת בינה מלאכותית יכולה להתבצע ברמות שונות, המחייבות דרגות שונות של מומחיות טכנית מצד המבקרים ודרגות משתנות של גישה לרכיבים הטכניים של המערכת:

1. ברמה הבסיסית (baseline), הביקורת מתמקדת בסקירת התיעוד כדי לוודא שיש התייחסות לכל רכיבי הממשל העיקריים, שהסיכונים מזוהים ושהותו אסטרטגיות מתאימות להפחתת הסיכונים. גישה זו יכולה להעריך את מידת השליטה של הארגון בגורמי הסיכון, אך היא לא קובעת אם מערכת הבינה המלאכותית מפירה בפועל עקרונות כמו יחס שוויוני.
2. כאשר המבקרים מקבלים גישה להרצת המודל ולבדיקתו באמצעות נתוני ייצור אמיתיים או נתונים סינתטיים, הם יכולים להעריך באופן ישיר את ההתנהגות ואת התוצרים של המערכת. זה מאפשר להם לבקר את השפעת המערכת בפועל. לדוגמה, המבקרים יכולים לבדוק מערכות חיזוי מבוססות בינה מלאכותית על ידי שינוי משתנה קלט יחיד והשוואת התוצאות, או לחפש פגיעויות במערכת בינה מלאכותית יוצרת על ידי שימוש בפרומפט שנוסחו במיוחד לצורך זה.
3. בביקורת מקיפה של מערכת בינה מלאכותית, בדיקות טכניות נוספות יכולות להקנות תובנות רבות ערך לגבי

⁵² למידע נוסף, ראו: [Directive 2019/790](#).

⁵³ למידע נוסף, ראו את הנחיות ממשלת בריטניה: [Copyright and Artificial Intelligence](#).

⁵⁴ EU AI Act, Article 95.

⁵⁵ למידע נוסף, ראו: [OECD, Measuring the Environmental Impacts of Artificial Intelligence Compute and Applications](#).

⁵⁶ ראו בנספח 1 תיאור של מבנה ביקורת בינה מלאכותית בהתאם למסגרת [CRISP-DM](#).

סקירה כללית של קטלוג הביקורת
נושאים עיקריים ושאלות ביקורת לדוגמה לפי תחום

4.1 ניהול וממשל של הפרויקט	4.2 נתונים
נושאים עיקריים • הצעת פרויקטים של בינה מלאכותית • דרישות חוקיות ואתיות בתחום השיפוט הספציפי • ניהול מחזור החיים של בינה מלאכותית • הערכת סיכונים של בינה מלאכותית • באילו תהליכים עסקיים ישמש היישום? • מה הן הראיות לניתוח עלות-תועלת?	נושאים עיקריים • כמות הנתונים ואיכותם • יצירה מבוססת-אחזור (RAG) ומודלי שפה גדולים (LLM) • סוגיות של פרטיות בהקשר של מערכות בינה מלאכותית מה הם מקורות הנתונים? כיצד תעריכו את איכות הנתונים?
4.3 כיתוח המודל	4.4 הערכה לפני פריסה
נושאים עיקריים • תהליך הפיתוח ומאפייני הביצוע • הבטחת איכות, מהימנות ועמידות בפיתוח • ציפיות לגבי תיעוד ובקרת גרסאות כיצד בחרתם את המאפיינים ששימשו ליצירת המודל? מה הם מאגרי הנתונים המשמשים אתכם לאימות (ולידציה)?	נושאים עיקריים • בדיקות ביצועים וקבלה • שקיפות והסברתיות • הוגנות ומזעור פגיעה • אבטחה • השפעה סביבתית כיצד אתם מבטיחים את הוגנות המודל? כיצד אתם מגינים על היישום מפני שימוש לרעה?
4.5 פריסה וניהול שינויים	4.6 מערכות בינה מלאכותית בסביבה תפעולית
נושאים עיקריים א. מוכנות עסקית וניהול שינויים ב. קריטריונים לעלייה לאוויר ושערים מה הם החלקים המרכיבים את מערכת למידת המכונה? כיצד היישום מוטבע בארכיטקטורת המערכת הסובבת?	נושאים עיקריים 1. ניטור ומעקב ביצועים 2. תגובה לתקריות ועדכוני המודל 3. ציות שוטף ונתיבי ביקורת כיצד אתם מנטרים את ביצועי המודל לאורך זמן? מה יניע אתכם לסקור את המודל או לבצע אימון מחדש?

כל תחום ביקורת כולל סיכונים ובקורות מצופות. השאלות לדוגמה נועדו להמחשה, ויש להתאימן לכל מערכת בינה מלאכותית העומדת לבחינה.

לוח 4.1: סקירה כללית של קטלוג הביקורת. השאלות לדוגמה נועדו להדגמת כלי העזר לביקורת.

מרכיבי המערכת:

- ביקורת מדוקדקת של הנתונים וסקירת הקוד יכולות להקנות ביטחון רב יותר לגבי מידת הדיוק של התיעוד, בפרט בנוגע למפרטים של מודל למידת מכונה.

- עבור מערכות חיצוניות מבוססות בינה מלאכותית ייתכן כי יידרש שחזור של האימון, הבדיקות, הדירוג ומדדי הביצועים של המודל (או חלקים ממנו) כדי להבין ולאמת את פרטי המודל, ביצועיו והדירותו, וכן את ההשלכות שלו לגבי הוגנות. הדבר דורש תשתית שמסוגלת לתמוך בתהליך אימות מסוג זה. יש להביא בחשבון את העלויות הקשורות בכך ואת התועלות הפוטנציאליות של גישה זו כאשר שוקלים לנקוט אותה.

- פיתוח חלופות מתאימות למודל יכול להועיל בהבלטת ליקויים ודרכים למנוע אותם. כמו בסעיף (ב3), נדרשת בחינה קפדנית בטרם מחליטים לפתח מודלים חלופיים, מפני שמהלך זה עלול להוביל לתפקידים החורגים מתפקידיו של מוסד הביקורת העליון הקבועים בחוק. כמו כן יש לשקול את העלות והתועלת של גישה זו. במקרה של מערכות בינה מלאכותית מורכבות ו/או מערכות שנרכשו מצד שלישי, יש להניח כי גישה כזו לא תהיה ישימה עבור המבקרים.

מבקרים המעוניינים לבצע את כל רמות הסקירה צריכים להיות בעלי ידע מעמיק על הטכנולוגיה של המערכת כדי לתת המלצות שיפור בעלות משמעות. כאשר הארגון המבוקר נעדר מומחיות פנימית, יש לזהות כל ליקוי ולהציע המלצות מתאימות. משימה זו עשויה להצריך מומחיות של אנשי מקצוע המתמחים בבינה מלאכותית או בטכנולוגיות מידע. עומק הביקורת המתאים ייקבע על פי הסיכונים הקשורים למערכת הבינה המלאכותית, מטרת הביקורת ושאלות הביקורת הכלליות⁵⁷.

גורם מגדיר נוסף בעת תכנון ביקורת בינה מלאכותית הוא בשלותו של הגוף המבוקר ושיאיותו ביחס לבינה מלאכותית, כמו גם השאלה האם המערכת פותחה על ידי הגוף המבוקר או על ידי גוף חיצוני שממנו נרכשה. רמת התיעוד הזמין ואופיו ישתנו משמעותית, קרוב לוודאי,

בהתאם למבנה הארגוני של הגוף המבוקר, השקעתו בתחום הבינה המלאכותית והיקף הפרייקט שלו. במקרה של מערכות הנרכשות מגורם חיצוני, כאשר קוד המקור ומפרט המודל המפורט אינם ידועים לגוף המבוקר (או שאין לו אפשרות לאמת אותם), הגוף המבוקר עדיין אחראי לספק תיעוד מקיף המוכיח את עמידתו בסטנדרטים נאותים של תהליכי מינהל ציבורי.

על המבקרים להניח כי הנתונים, או מקטעי נתונים רלוונטיים, זמינים לצד התיעוד. עם זאת ייתכן שהקוד והמודל עצמו אינם בפורמט נגיש למבקרים⁵⁸, במיוחד במקרה של מערכות בינה מלאכותית שנרכשו מגוף חיצוני. במקרים כאלה, שיתוף פעולה הדוק עם הארגון המבוקר הוא הכרחי לצורך בדיקות ו/או הדגמות באתר הארגון, שיבוצעו בפיקוחם של מבקרים בעלי הבנה נאותה בנושא.

4.1 ניהול וממשל של הפרייקט

ניהול וממשל מועילים של הפרייקט הם יסוד חיוני להצלחת הפיתוח והפריסה של מערכות בינה מלאכותית במגזר הציבורי.

ביקורת של ניהול וממשל בפרייקט בינה מלאכותית מתמקדת בפיקוח הארגוני, ולא בביצועים הטכניים של המערכת. היא מבוצעת ברמת הממשל ומתבססת בראש ובראשונה על סקירת מסמכים וראיונות, ואינה כרוכה בבדיקות ישירות של מערכת הבינה המלאכותית.

השלב הראשון בתהליך הוא בחינת הגודל והמבנה של הארגון הנבדק, שכן גורמים אלה משפיעים על הסדרי הממשל ועל היקף התיעוד הזמין. בארגונים גדולים ייתכן שכבר גובשו אסטרטגיות ועקרונות לתחום הבינה המלאכותית, ואילו בארגונים קטנים ייתכן שהפרייקט מנוהל באופן פחות רשמי. כאשר היקף התיעוד מוגבל, ראיונות יכולים להקנות תובנות חשובות לגבי פרקטיקות ניהול הפרייקט. לכל הפחות, על המבקרים לצפות לראות ראיות למרכיבי הממשל העיקריים, כפי שמתואר בתתי-הסעיפים הרלוונטיים שכותרתם "בקורת מצופות".

4.1.1 הצעת פרייקטים בתחום הבינה המלאכותית

הבנת מקורותיו ומטרותיו של פרייקט בינה מלאכותית היא צעד ראשון קריטי. מרבית הארגונים דורשים שהצעה לפרייקט תאושר לפני תחילת העבודה, אם כי רמת

57 לדוגמה, במסמך "שימוש בבינה מלאכותית בממשלה המרכזית" (The Use of Artificial Intelligence in the Central Government), שאלות הביקורת הכלליות נגעו לשימוש אחראי בבינה מלאכותית בממשלה המרכזית, ואילו הפרטים הטכניים של המערכות שהיו בשימוש בעת ההיא לא היו חלק מתכולת הביקורת. כך שהביקורת של ארבע מערכות בינה מלאכותית שהוצגו כחקרי מקרה במסמך בוצעה בפורמט של ביקורת בסיסית (baseline) שבדקה היבטים של ממשל בינה מלאכותית.

58 מודלי למידת מכונה רבים מפותחים בשפת פייתון או R, אבל יש גם אפשרויות אחרות לכתיבת התוכנה. לא סביר אפוא לדרוש שהמבקרים יכירו כל שפת קוד ומסגרת תוכנה שקיימות. העלויות הקשורות לרכישת תוכנה נוספת ו/או כוח מחשוב נוסף עלולות ליצור בעיות נוספות.

הפורמליות עשויה להשתנות. הצעה נאותה תכלול פרטים אלה:

- המטרות והיעדים של הפרויקט.
- ההיקף המיועד והתוצרים.
- מדדי ביצועים וקריטריונים להצלחה.
- הערכה של דרישות מוקדמות, כמו זמינות נתונים ומומחיות רלוונטית.
- ניתוח עלות-תועלת.
- זיהוי בעלי עניין מרכזיים.
- הערכה של טכנולוגיות מתאימות.
- הסדרים לניהול מערכת הבינה המלאכותית לכל אורך מחזור החיים שלה.

הערכת דרישות מוקדמות כמו נתונים זמינים ומומחיות נדרשת, וגיוש הבנה ברורה של הפוטנציאל והמגבלות של טכנולוגיית הבינה המלאכותית, דורשים על פי רוב שיתוף פעולה בין מומחים טכניים למומחים מקצועיים לתחום העיסוק (domain experts) בשלבים המוקדמים ביותר.

אפילו בפרויקטים גמישים, שהתוכניות לגביהם עשויות להשתנות, חשוב להגדיר מטרות, דרישות ומדדי ביצוע בשלבים הראשונים.

כאשר קיימים כללי מדיניות פנימיים שמתייחסים לבינה מלאכותית, תוכניות הפרויקט צריכות להפנות אליהם כדי להבטיח תיאום עם עקרונות הממשל הארגוניים - גם עבור מערכות שנרכשו או פותחו מחוץ לארגון.

עבור פרויקטים מורכבים או בעלי עלות גבוהה, הוכחת היתכנות (proof of concept) יכולה לסייע לבדוק את הישומות ולהקטין את הסיכון לתוצאות בלתי מספקות. בתהליך זה יש להעריך הן את הביצועים הטכניים, הן את היכולת לעמוד בדרישות פונקציונליות. פרויקטים של בינה מלאכותית מעל היקף מסוים עשויים להוביל לדרישות רגולטוריות נוספות בנוגע למחקרי ישימות ותיעוד.

מערכת בינה מלאכותית צריכה להראות יתרונות ברורים ומדידים בהשוואה לגישות קונונציונליות כדי להצדיק את מורכבותה, עלותה והסיכונים הכרוכים בה. אף שמודלי למידה עמוקה נפוצים בשימוש, הם גוררים לעיתים קרובות עלויות שאינן נראות לעין מיד. לדוגמה, אימון או כוונן עדין של מודלי למידה עמוקה מצריכים שימוש בחומרה מתמחה, כגון יחידות עיבוד גרפיקה (GPU), שניתן לפרוס בארגון באופן מקומי או לשכור מפלטפורמות ענן. פריסה מקומית בארגון מצריכה השקעה משמעותית בחומרה, באנרגיה ובכירור, ונלווה לה הסיכון הנוסף של התיישנות עם הזמן. פריסה מבוססת ענן מתומחרת בדרך כלל במודל של תשלום לפי שימוש (pay-as-you-go), כך שהגבירה היא לפי עומס העבודה הנדרש לאימון והרצה של המודל. העלויות עלולות לעלות במהירות בהתאם למשך הזמן

והאינטנסיביות של תהליכים אלה.

למרות אתגרים אלה, מערכות בינה מלאכותית עשויות לספק יתרונות מהותיים. ביניהם אפשר לציין שיפורי יעילות ופרודוקטיביות באמצעות מיכון מטלות, הפחתת טעויות אנוש והגברת החדשנות. מודלי למידת מכונה, לדוגמה, מסוגלים לזהות דפוסים וקשרים שאינם ניכרים לעין האנושית בנקל, כך שהם מאיצים תהליכים כמו גילוי תרופות או ניתוח נתונים בקנה מידה גדול. בינה מלאכותית יכולה גם לחזק אבטחת סייבר ולתרום לניצול מיטבי של מערכות אנרגיה, כולל תכנון תחנות כוח ואמינות אספקת החשמל. במגזר הציבורי, בינה מלאכותית מציעה הזדמנויות לניהול יעיל ושירותים מותאמים אישית בעלות נמוכה יותר.

על המבקרים לוודא כי בוצע ניתוח עלות-תועלת יסודי לפני הטמעת הפרויקט, וזאת בפרט במקרה של מודלי יסוד ו-LLM, שבהם תמחור המבוסס על צריכה עלול לגרום להוצאות שוטפות בהיקף משמעותי. ניתוח עלות-תועלת יכול:

- **הגדרת היעדים** שהגוף המבוקר חותר להשיג באמצעות מערכת הבינה המלאכותית, ומדדי ביצוע עיקריים (KPI) הקשורים לאותם יעדים.
- **הערכה השוואתית** כנגד מערכות קיימות וחלופות ישימות, המביאה בחשבון ביצועים, עלויות והשפעה עסקית. הערכה זו צריכה לכלול ניתוח של רמת הדיוק של המודל ועלויות הקשורות לטעויות, הוצאות פיתוח ותחזוקה וחיסכון פוטנציאלי. כדי לקבוע שהמודל מפיק תוצאות מהימנות, יש ליישם שיטות בדיקה מעמיקות כמו השוואות נתונים היסטוריות, מחקרי פיילוט או בדיקות A/B.
- **תרגום ביצועים טכניים לערך רלוונטי בהיבט העסקי** (פיננסי, תפעולי, אסטרטגי).
- **זיהוי עלויות נסתרות ולאורך מחזור החיים**, כולל השפעה סביבתית.
- **ניתוח עלויות טוקנים/API** עבור מודלי יסוד, והשפעתן על כדאיות בטווח הארוך.
- **מידת שיפורי פרודוקטיביות** וסימנים ל"בינה מלאכותית לשם עצמה".
- **תחזיות של עלות התחזוקה לאורך מחזור החיים**, כולל שיקולים של סוף החיים.
- **הערכת עלויות משפטיות, עלויות ציות ועלויות הזדמנות.**

4.1.1.1 סיכונים שיש להביא בחשבון

- אימוץ פתרונות בינה מלאכותית ללא ערך מוסף ברור או התאמה מובהקת.
- הערכה לא נאותה של גישות חלופיות, כמו שיטות אנליטיות או מבוססות כללים.
- הבנה טכנית מוגבלת בקרב אנשי ההנהלה, או מודעות בלתי מספקת למציאות התפעולית בקרב הצוותים הטכניים, שעלולות לגרום לתקשורת לא טובה ולציפיות בלתי מציאותיות.
- היעדר אימות של ישימות המערכת באמצעות הוכחת היתכנות.
- היעדר הגדרה של מטרות או מדדי ביצוע.
- חוסר התאמה לעקרונות הפנימיים של ממשל בינה מלאכותית.
- עלויות מחשוב מופרזות, כמו חומרה, אנרגייה או חיובי ענן, עקב שימוש במודלים מורכבים שלא לצורך.
- עלויות טוקן או API גבוהות עקב שימוש ב-LLM היכן שמודלים כשוטים יותר יספיקו.

4.1.1.2 בקורות מצופות

- מעורבות מוקדמת ושוטפת של בעלי עניין. לשם כך אפשר להיעזר במנגנונים כמו סדנאות בשיתוף גורמים שונים, אסיפות אזרחים או התייעצויות עם הציבור.
- הצהרות מוגדרות כראוי של בעיות והנחות מוסכמות.
- פיתוח משותף של מטרות פרויקט הבינה המלאכותית, ותוכנית פרויקט הכוללת את הבעלים של הפרויקט ובעלי עניין רלוונטיים, כדי להבטיח ישימות והבנה משותפת.
- יכולת לקשר באופן ברור בין המטרות לדרישות הטכניות והפונקציונליות (traceability), אשר מתועדות כראוי ומתורגמות לדרישות מודל המאפשרות מדידה וניטור.
- הגדרת אינדיקטורים למדידת הערך המוסף של מערכת הבינה המלאכותית מראש, המבטיחים יכולת להעריך את היתרונות לכל אורך מחזור החיים.

- תיעוד של ניתוח עלות-תועלת, ניתוח סיכונים והמלצות לגבי הטכנולוגיה.
- הוכחת היתכנות במקרה של פרויקטים גדולים.
- התייחסות להשפעה הסביבתית לפני פריסה.
- מיפוי פרויקט הבינה המלאכותית אל מול אסטרטגיית הבינה המלאכותית של המוסד ו/או הנחיותיו בנושא, במקרה הרלוונטי.

4.1.2 דרישות חוקיות ואתיות החלות בתחום השיפוט הספציפי

פרויקטים של בינה מלאכותית כרוכים לרוב בטווח רחב יותר של שיקולים משפטיים ואתיים לעומת מערכות IT מסורתיות, בפרט כאשר מדובר במערכות מורכבות, מערכות אוטונומיות או כאלה המשפיעות באופן ישיר על בני אדם. הרלוונטיות של תקנות ספציפיות תלויה בגורמים כמו שימוש בנתונים אישיים, רמת האוטומציה, המשתמשים המיועדים ומטרות המערכת. לדוגמה, מערכות בינה מלאכותית שיעודן רק לתמוך בתהליכים פנימיים, ואינן משפיעות על אזרחים ישירות, יצריכו בדיקת ביקורת מעמיקה פחות בסוגיות של אתיקה בהשוואה למערכות קבלת החלטות אוטומטית, שיש להן השפעה ישירה או עקיפה על אזרחים.

אומנם תקנות אסדרה וחוקים רבים אינם מתייחסים לטכנולוגיה ספציפית, אך תרגום עקרונות כמו מניעת אפליה ופרטיות לדרישות טכניות עלול להיות מאתגר. הסטנדרטים לפירוש מושגים משפטיים - כמו אפליה בהקשר של אלגוריתמים - ולתרגום שלהם לדרישות טכניות, נמצאים עדיין בתהליך של התפתחות⁵⁹. על המבקרים לבדוק ולאשר שצוות הפרויקט כולל אנשים בעלי מומחיות משפטית מספקת, ושהקווים המנחים הפנימיים בארגון מתייחסים לאתיקה ולסיכונים המשפטיים. הערכת החוקים והתקנות הרלוונטיים תכלול הן תפעול רגיל, הן תופעות לוואי פוטנציאליות, לרבות כאלה הנובעות מפגמים או תקלות במודל.

אתגרים אלה מצטמצמים במידה מסוימת בזכות גיבושן ויישומן של תקנות אסדרה ייעודיות לבינה מלאכותית בתחומי שיפוט רבים (ראו חלק 3). ה-EU AI Act, לדוגמה, בהיותו אסדרה של בטיחות המוצר בהגדרתו, מכיל יותר פרטים טכניים מחקיקה שמתמקדת בזכויות יסוד. כך הוא מקשר את דרישות החוק למפרטים טכניים בהתבסס על מאפייני מערכת הבינה המלאכותית. אף שהדבר יכול לסייע

59 סקירה מקיפה של ההשלכות הפרקטיות של עקרונות בינה מלאכותית, כולל המקרה של בינה מלאכותית יוצרת ותיאור פערי מחקר, ניתן למצוא כאן: A. Oesterling et al., [Operationalizing the Blueprint for an AI Bill of Rights: Recommendations for Practitioners, Researchers, and Policy Makers \(2024\)](#).

- קווים מנחים פנימיים בדבר שימוש אתי בבינה מלאכותית ובדבר סיכונים משפטיים.
- הערכה מקיפה של החוקים והתקנות הרלוונטיים.
- ניתוח השלכות משפטיות ואתיות של שימוש בשירותים או מודלים חיצוניים.

4.1.3 ניהול מחזור החיים של בינה מלאכותית

ניהול מקיף של מחזור החיים יבטיח שפרויקטים של בינה מלאכותית יהיו מתוכננים כראוי וייתמכו משלב ההקמה ועד סיום חיי המערכת, ובכלל זה:

- הגדרה של המטרה, הגבולות והמגבלות של יישום הבינה המלאכותית.
- תכנון לקראת שלבי העיצוב, הפיתוח, אבטחת האיכות, הפריסה, התחזוקה, העדכונים והוצאת המערכת משירות פעיל בסיום חייה.
- הטמעת עקרונות אתיים - כמו הגנת מידע, שקיפות ואי-אפליה - בכל שלבי מחזור החיים ("Ethical AI by Design").
- עבור מערכות שנרכשות או מתקבלות דרך מיקור חוץ, יש להבטיח שניהול מחזור החיים יכסה את הרכש, ניהול החוזה ותיאום עם עקרונות ממשל מקומיים.

4.1.3.1 סיכונים שיש להביא בחשבון

- חוסר בהירות של התפקידים ותחומי האחריות בארגון המבוקר עבור כל שלבי הפרויקט, המוביל לעיכובים או לפגיעה בפונקציונליות.
 - הגנה בלתי מספקת על עקרונות האתיקה, אם אלה אינם מוטבעים בתכנון עצמו לאורך מחזור החיים הכולל של המערכת.
 - חוסר יכולת לתמוך במערכת הבינה המלאכותית לאורך כלל מחזור החיים שלה, כולל תחזוקה, עדכונים וניהול תקריות.
 - שקיפות ושליטה מוגבלות כאשר הפיתוח, התשתית או רכיבים של המערכת מתנהלים במיקור חוץ, מצב המוביל לבעיות הבאות:
1. מידע שאין די בו כדי לקבל החלטות מושכלות או לבקר את המערכת;
 2. תפקידים ותחומי אחריות לא ברורים, המגדילים את הסיכונים ופוגעים ביכולת להגיב לתקריות או לנתח כשלים;

למפתחים להטמיע עקרונות אתיים בפועל, הוא גם דורש הערכה קפדנית של תחולת האסדרה. ייתכן שהמפתחים ינסו להימנע מחובות החוק על ידי ניצול מצבים של חוסר בהירות או פטורים בתחולת האסדרה - כמו כללי סיווג, רמות סיכון או קטגוריות של מפעילים (משתמשי המערכת).

כאשר ארגונים משתמשים בשירותים או מודלים חיצוניים, עליהם להעריך את ההשלכות במישור המשפטי והאתי, לרבות ציות לדינים מקומיים ואותנטיות של נתוני האימון. ייתכן מצב שבו בפיתוחם של מודלים חיצוניים הופרו תקנות מקומיות, או שהפיתוח עורר חששות אתיים כמו הפרות פוטנציאליות של זכויות קניין רוחני במהלך האימון של LLM. על הארגונים לשקול גם את הסיכונים לטווח ארוך של הסתמכותם על ספקים זרים או על מודלים שפותחו מחוץ לארגון. כמו כן תקנות האסדרה עשויות להגביל את מיקומם הפיזי של שירותים מתארחים, ומגבלות מסוג זה יכולות להשתנות על פני זמן - כפי שהודגם בשינויי החקיקה להעברות נתונים בין האיחוד האירופי לארה"ב בעקבות פסקי הדין בעניין (Schrems I, Schrems II)⁶⁰.

4.1.2.1 סיכונים שיש להביא בחשבון

- ידע בלתי מספיק או חוסר תשומת לב להוראות של חוקים וכללי אתיקה.
- הפרת דיני הגנת המידע או חוקים נגד אפליה.
- הערכה מאוחרת או חסרה של מסגרות אסדרה, שעלולה להוביל לעיכובים בפרויקט, לעלויות מוגדלות או לעצירת פרויקטים.
- עקיפת דרישות האסדרה באמצעות טקטיקות כמו המעטה בחשיבות פונקציונליות מערכת הבינה המלאכותית או הגבלתה באופן מלאכותי, תיוג שגוי של מערכת הבינה המלאכותית או סוג הפרויקט, או טכניקות דומות שמטרתן התחמקות מציות.
- חוסר מודעות לסיכונים משפטיים או אתיים הקשורים לפיתוח מודלים חיצוניים או למיקומם הפיזי של שירותים חיצוניים.
- אי-ודאות לגבי הישימות ארוכת הטווח של שימוש בשירותי ספקים זרים (בחו"ל).

4.1.2.2 בקורות מצופות

- צוותי פרויקט רב-תחומיים בעלי מומחיות משפטית, מומחיות טכנית ומומחיות בתחום העיסוק (domain experts).

60. Reuters: [EU seals new US data transfer pact, but challenge likely](#)

3. בעיות של המשכיות ביצועים בסביבות המשתנות במהירות, כאשר המודלים הבסיסיים או הפלטפורמות מתעדכנים לעיתים קרובות ללא שליטה של הגוף המבוקר.

- רכישת מערכות או שימוש במערכות שפותחו מחוץ לארגון ללא תמיכה לטווח הארוך, בשילוב עם היעדר מומחיות פנימית בארגון, העלולים להוביל לגריעת ביצועים ולהגברת סיכוני האבטחה.

4.1.3.2 בקורות מצופות

- תוכנית מחזור חיים מתועדת, לרבות תפקידים⁶¹, תחומי אחריות, צירי זמן ומשאבים.
- עקרונות אתיקה המובנים מראש בתכנון (by design).
- במקרה של מערכת נרכשת או כזו שחלק מהפיתוח שלה בוצע במיקור חוץ, יש לנקוט אמצעים כדי לוודא שהגוף המבוקר שומר על שליטה מספקת על השימוש בבינה מלאכותית, לרבות:

1. תיעוד מקיף המגדיר בבירור את התפקידים ותחומי האחריות - הן של הגוף המבוקר, הן של הספק החיצוני;
2. במקרים רלוונטיים, הסכמי רמת שירות (SLA) המבטיחים תמיכה שוטפת ומספקים בקרה נאותה על תשתיות חיצוניות או שירותים חיצוניים, כדי להבטיח ניהול תקריות וניתוח כשלים יעיל.
- תוכנית הבטחת איכות מוסכמת המתואמת עם מטרות הפרויקט, והדרישות הטכניות והפונקציונליות הקשורות בכך.

4.1.4 הערכת סיכונים

מערכות בינה מלאכותית מייצרות סיכונים אשר לא בהכרח מקבלים מענה במסגרות ניהול הסיכונים המקובלות. ככל שהטכנולוגיה הולכת ומתפתחת, כך מתפתחת גם ההבנה שלנו לגבי הסיכונים הטמונים בה. רבים מסיכונים אלה נובעים מהדרך שבה מערכות בינה מלאכותית לומדות מנתונים, וכן ממגבלות השקיפות והשליטה על הדפוסים שהן מזהות. לדוגמה, בינה מלאכותית עלולה לחזק הטיות

קיימות שלא במתכוון, לחשוף מידע סודי המוטמע בנתוני האימון, או להיחשף למניפולציות באמצעות טכניקות הטעיה. סיכונים אלה גוברים כאשר ארגונים משתמשים במודלי יסוד שפותחו מחוץ לארגון, ושלא תמיד ברור מה היו נתוני האימון שלהם ומה הם אמצעי ההגנה.

כאשר מבצעים ביקורת כדי לבדוק את הניהול של בינה מלאכותית בגוף המבוקר, יש לבדוק אם מסגרת ניהול הסיכונים של הארגון עודכנה כדי לשקף חששות הנוגעים לבינה מלאכותית באופן ספציפי. ברמת הפרויקט, אופיים וחומרתם של סיכוני הבינה המלאכותית ישתנו בהתאם ליישום. עם התחומים העיקריים הראויים לתשומת לב המבקרים נמנים תחומי האבטחה, ההוגנות, השקיפות וההסבריות - כל אחד מהם נבחן ביתר פירוט בתתי-הסעיפים הרלוונטיים של סעיף 4.4.

אחת השאלות המרכזיות שעל המבקרים לבדוק היא האם הארגון ערך ניתוח סיכונים המתייחס מפורשות לסיכונים הייחודיים של בינה מלאכותית. ניתוח זה צריך להתייחס לא רק לסיכוני הפרויקט, אלא גם להשפעותיו הפוטנציאליות על פרטים ועל החברה.

4.1.4.1 סיכונים שיש להביא בחשבון

- הערכות סיכונים סטנדרטיות עלולות להתעלם מסיכונים ייחודיים של בינה מלאכותית.
- סיכונים תפעוליים, טכניים וחברתיים הנובעים מבינה מלאכותית לא תמיד מזוהים באופן מלא, לא מוסברים, או שלא ננקטים אמצעים להפחתתם.

4.1.3.2 בקורות מצופות

- במקרים הרלוונטיים, הערכת שיקולי אתיקה נוסף על סיכוני פרויקט מוכרים וסיכוני בטיחות.
- הכללת סיכונים הקשורים להסתמכות על תשתיות חיצוניות או מודלים חיצוניים.
- כאשר משתמשים במודלים שפותחו מחוץ לארגון או בקוד פתוח, הבאה בחשבון של סיכוני בטיחות הנובעים מגורמים זדוניים, כגון החדרת דלתות אחוריות (backdoors) או בעיות בהתאמת המודל למטרותיו.
- התייחסות למסגרות מבוססות לניהול סיכוני בינה

61 ראו בנספח 5 רשימה של תפקידים אפשריים.

והצורך בניטור ותחזוקה שוטפים. פרק זה מתאר את הסיכונים העיקריים הקשורים בנתונים ואת הבקורות הרלוונטיות שיש לבחון במסגרת ביקורת מערכת בינה מלאכותית.

4.2.1 כמות הנתונים ואיכותם

מהימנותה של כל מערכת בינה מלאכותית תלויה באיכות נתוני האימון שלה ובהיותם מספיקים (דיות נתונים). נתונים בלתי הולמים או מוטים עלולים להוביל למודל שמייצר תוצאות בלתי הוגנות ולא מדויקות. לדוגמה, אם נתוני האימון אינם משקפים את אוכלוסיית העולם האמיתי, המודל עלול להעדיף באופן שיטתי קבוצות מסוימות על פני אחרות.

בעיות שכיחות אחרות באיכות הנתונים במערכות בינה מלאכותית הן תוצר לא מדויק או לא שלם - טעויות, השמטות או מידע שעבר זמנו עלולים לערער את ביצועי המודל. חוסר עקביות עלול להיווצר כאשר מצרפים נתונים ממקורות שונים, דבר העלול להוביל לאי-התאמות של הפורמט או המשמעות.

במקרה של מערכות בינה מלאכותית שפותחו על ידי גוף חיצוני או נרכשו ממנו, ייתכן שלארגון תהיה גישה מוגבלת לנתוני האימון, והדבר יקשה להעריך את התאמתם להקשר שבו פועל הארגון.

נתונים בלתי מספיקים מייצגים אתגר משמעותי בפיתוח מערכות של בינה מלאכותית. כמות הנתונים הנדרשת גדלה יחד עם מורכבות הבעיה שהמערכת מיועדת לטפל בה. כאשר הנתונים הזמינים מוגבלים מדי יחסית למורכבות המודל, קיים סיכון של התאמת יתר (overfitting). זהו מצב שבו המודל לומד את נתוני האימון - כולל הרעש שבהם - באופן מדויק מדי, ואינו מצליח להפיק מהם הכללות לגבי מצבים חדשים. במצב ההפוך, אם המודל פשוט מדי למשימה, הוא עלול להימצא בהתאמת חסר (underfitting). אז הוא יחמיץ דפוסים חשובים או לא יצליח לתפוס את המבנה שמאחורי הנתונים.

בלמידה מונחית (supervised learning) דרושים נתונים מספיקים כדי לבצע אימון, אימות ובדיקות. נתונים בלתי מספיקים בכל אחד משלבים אלה, או שימוש באותם נתונים לשלבי האימון והבדיקות גם יחד, יובילו למדדי ביצוע

מלאכותית⁶² וקווים מנחים להערכות הנדרשות על פי חוק⁶³, והסתייגות בהם כדי לזהות סיכונים בינה מלאכותית ייעודיים ולטפל בהם, גם כאשר הדבר אינו מתחייב מהוראות החוק.

- פיתוח מודיעין אסטרטגי כדי לצפות מראש שיבושים ולנהל סיכונים מתפתחים, באמצעות כלים כמו סריקת אופק וניתוח תרחישים⁶⁴.
- בחינה והשוואה של הסיכונים הכרוכים בגישות חלופיות בשלב של הצעת הפרויקט ובחירת המודל (ראו סעיפים 4.11 ו-4.3).
- מיפוי שיטתי של נזקים פוטנציאליים כאשר מדובר בתחומי פעולה קריטיים (כמו שירותי בריאות או אכיפת החוק), ושימוש בסולמות מדידה מתוקננים להערכת חומרת ההשפעות, לרבות אלה הקשורות למידע מוטעה, אבטחה, אפליה והפרעה חברתית-כלכלית.
- סקירת תיעוד לצורך הערכת סיכונים:
 1. סקירת התיעוד הטכני, כולל שפות קוד, גרסאות של פלטפורמות, תוצרי המודל והיסטוריה של הפרומפטים למודל והגדרות יצירת התוכן עבור בינה מלאכותית יוצרת.
 2. בדיקת הציות לדיני הפרטיות, להנחיות בנושא אתיקה ולמדיניות הפנימית לבינה מלאכותית.
 3. הערכת בקורות איכות, לרבות תיעוד תהליכי MLOps/LLMOps, פעילויות צוות אדום (המחפש נקודות תורפה) וניטור לוחות מדדים (דשבורד) (ראו סעיף 4.3).
 4. וידוא קיומו של תיעוד מפורש לכל הנחות היסוד שעליהן התבססו הערכות המערכת.

4.2 נתונים

האפקטיביות של מערכות בינה מלאכותית תלויה בנתונים שעליהן הן בנויות. אף שממדי איכות נתונים מקובלים - כמו שלמות, תקפות וייחודיות - עדיין חשובים, הבינה המלאכותית מייצרת אתגרים נוספים. עם אלה נמנים האיכות, הכמות והייצוגיות של נתוני האימון, וכן הסיכונים הקשורים להטיה, התאמת יתר (overfitting - ראו להלן)

62 לדוגמה: National Institute of Standards and Technology. (2023), [Artificial Intelligence Risk Management Framework \(AI RMF\)](#) או the IBM AI Risk Atlas: F. Bagehorn et al., [AI Risk Atlas: Taxonomy and Tooling for Navigating AI Risks and Resources \(2025\)](#).

63 לדוגמה, קיימים קווים מנחים המתארים מתי וכיצד לבצע הערכת השפעות על הגנת המידע (כנדרש על פי GDPR באיחוד האירופי), ברמה בין-לאומית וברמה לאומית. אנו מצפים כי יתפרסמו קווים מנחים דומים לגבי הערכת השפעות על זכויות יסוד, כנדרש על פי ה-EU AI Act במקרים ספציפיים.

64 OECD (2024), ["Framework for Anticipatory Governance of Emerging Technologies"](#), OECD Science, Technology and Industry Policy 64 Papers, No. 165, OECD Publishing, Paris.

- הגדרת המאפיינים העיקריים של האוכלוסייה ובדיקה אם הם מיוצגים כראוי בנתוני האימון, הבדיקה והתיקוף.
- סקירה ועדכון סדורים של נתוני האימון השמורים כדי למנוע התיישנות שלהם.
- עבור מערכות שפותחו על ידי גוף חיצוני או נרכשו ממנו - השגת מידע מספיק לגבי נתוני האימון מהגורם המפתח, אימות ביצועי המערכת עם נתונים מקומיים ובדיקה לגילוי הטיות פוטנציאליות.

4.2.2 מודלי שפה גדולים (LLM) ויצירה מבוססת אחזור (RAG)

יצירה מבוססת אחזור (retrieval-augmented generation - RAG) היא גישה המאפשרת למודלי שפה גדולים לשאוב ממקורות מידע חיצוניים בזמן שהם יוצרים תשובות. הזנת מסמכים או נתונים רלוונטיים לתוך ההקשר של המודל מאפשרת ל-RAG להתמודד עם אתגרים ידועים בבינה מלאכותית יוצרת, כמו הזיות, מידע מיושן והיעדר ציטוטי מקורות. עם זאת טכניקה זו מכניסה סיכונים חדשים הנוגעים לאיכות הנתונים המאוחזרים ואמינותם, וכן בקשר לבנייה ותחזוקה של בסיס הידע של המערכת⁶⁵.

מידע על הנתונים ששימשו לאימון מודלי יסוד שפיתחו ארגונים חיצוניים בדרך כלל אינו זמין. נתונים הנאספים אוטומטית ממקורות פתוחים (לדוגמה, מהאינטרנט) עלולים להיות פגיעים להרעלת נתונים. אם נתונים מסוג זה משמשים לאימון, לכוונן עדין או במערכות בינה מלאכותית שעלו לאוויר (לדוגמה, סוכני בינה מלאכותית המשתמשים בחיפוש באינטרנט), על המבקרים לצפות לראות ראיות לכך שהסיכון להרעלת נתונים הובא בחשבון ונוהל. סיכון זה רלוונטי גם ל-RAG, היכן שבסיס הידע נבנה על ידי שימוש בנתונים הנאספים באופן אוטומטי.

לשם בדיקה והערכה של פלטי LLM - כולל אלה המופקים באמצעות RAG - נדרשות לעיתים קרובות כמויות גדולות של נתוני דוגמאות המשקפים את מגוון שאילתות המשתמש האפשריות ואת התשובות הצפויות. בפועל, איסוף ידני של מאגרי נתונים כאלה כמעט אף פעם אינו ישים, בפרט ביישומים מורכבים או מתמחים. גישה חלופית היא לייצר נתונים סינתטיים - הנוצרים באופן מלאכותי במקום להיאסף מאינטראקציות בעולם האמיתי - ולהשתמש בהם כדי ליצור מערכי בדיקה מגוונים וסקלבליים לצורכי אימון, כוונן עדין או אימות של המודל. למרות יתרונותיהם, נתונים סינתטיים מצריכים אימות קפדני כדי לוודא שהם

מנופחים. טכניקות כמו הגדלת מספר הדגימות מקטגוריות פחות שכיחות (upsampling) או הקטנת מספר הדגימות מקטגוריות נפוצות יותר (downsampling) יכולות לסייע לשמור על איזון וייצוגיות בנתונים. סיכון נוסף מכונה דליפת יעד (target leakage): מצב שבו נתוני האימון כוללים בשוגג מידע שלא יהיה זמין בזמן החיזוי בפועל. הדבר עלול להוביל לביצועים שנראים מרשימים במיוחד בשלב הפיתוח, אך לתוצאות חלשות בתנאי אמת. דליפת יעד נובעת ממשנתנים הכוללים בטעות מידע המרמז על התוצאה הצפויה.

מודלים של בינה מלאכותית פגיעים גם לסיכונים כמו הרעלת נתונים, מצב שבו גורמים זדוניים מחדירים נתונים פגומים למאגר נתוני האימון כדי לערער את תקינות המודל ואמינותו.

גם תשתית הנתונים חשובה. מבני סילו של נתונים (data silos) - מצב של פיצול מידע בין מערכות שונות - יכולים לפגום ביכולת לבצע שילוב וניתוח של נתונים. ככל שנפח הנתונים גדל, נדרשת תשתית ניתנת להרחבה לצורך תמיכה בפיתוח ובתפעול השוטף. ההתייחסות לשיקולי אבטחה ופרטיות חיונית, במיוחד כשמטפלים במידע רגיש או אישי.

4.2.1.1 סיכונים שיש להביא בחשבון

- שימוש בנתונים בלתי אמינים, מוטים או בלתי מייצגים.
- הכנסת הטיה במהלך עיבוד הנתונים.
- הפרדה בלתי מספקת בין נתוני האימון לנתוני הבדיקה.
- יכולת הכללה גרועה של נתונים חדשים.
- הרעלת נתונים או מניפולציה לצורך הטעיה.
- דליפת יעד.
- היעדר שקיפות לגבי ייצוגיות נתוני האימון במודלים המפותחים מחוץ לארגון.

4.2.1.2 בקורות מצופות

- הערכת האותנטיות של מאגרי הנתונים ששימשו לאימון וכוונן עדין שלהם, בפרט כשמשתמשים בנתונים מהאינטרנט או נתונים סינתטיים, כדי להבטיח ציות לדרישות קניין רוחני והגנת המידע.
- וידוא שפיצולי הנתונים לצורכי אימון, תיקוף ובדיקה שומרים על ייצוגיות ומטופלים כראוי.
- השוואה של איכות הנתונים ואיכות המודל בכל מערכות הנתונים - לצורכי אימון, בדיקה ותיקוף.

65 פרטים טכניים נוספים על ארכיטקטורת RAG והערכת תוצרי RAG מובאים בסעיפים 4.3 ו-4.4.

הגנת המידע המתאימות.

4.2.3.1 סיכונים שיש להביא בחשבון

- הפרות של תקנות הגנת המידע (ראו נספח 2).

4.2.3.2 בקורות מצופות

- הערכת הסיכונים הנוגעים לנתונים אישיים, המסייעת לקבוע אם נדרש DPIA.
- DPIA, במקרה הרלוונטי.

4.3 פיתוח המודל

פיתוח המודל הוא שלב מכריע במחזור החיים של מערכת בינה מלאכותית, שבו העיצובים המושגים מתורגמים למודל תפעולי. עבור המבקרים, שלב זה מספק את התמונה הברורה ביותר של אופן הטמעתן של דרישות טכניות, ארגוניות ואתיות במערכת.

סעיף זה מסביר את התהליך של בניית מודל והבטחת איכותו בשלב הפיתוח, וכולל את הנושאים של בחירת מודל והתאמתו (כולל שימוש במודלי יסוד LLM), הנדסת מאפיינים, תכנון פרומפטים ופרמטרים, הבטחת איכות פנימית ושיטות להבטחת אמינות, עמידות והדירות. אנו מתייחסים לבדיקות קבלה (acceptance testing) מבוססות תוצרים ולהחלטות אור ירוק/אדום בסעיף 4.5, וסעיף 4.6 עוסק בבדיקות תפעוליות וסביבות חיות.

4.3.1 תהליך הפיתוח ומאפייני הביצוע

פיתוח של מודל בינה מלאכותית כרוך בבנייה, אימון ושכלול של המודל לצורך ביצוע משימות ספציפיות. תהליך זה כולל על פי רוב הכנת נתונים (ראו סעיף 4.2), בחירה או התאמה של אלגוריתמים, הערכת ביצועים ואופטימיזציה של הפרמטרים כדי לוודא שהמודל יספק ביצועים מדויקים כאשר יוזנו אליו נתונים חדשים. כיום הנטייה בארגונים היא להתאים מודלים שאומנו מראש או מודלי יסוד, כמו LLM, במקום לפתח מודלים מאפס. על המבקרים להעריך אפוא את מקור המודל, הרישוי, שיטות ההתאמה (כמו כוונן עדין, הנדסת פרומפטים או RAG) וכל קשרי התלות ההדדית עם ספקים חיצוניים.

בחירת המודל הנכון היא קריטית, מכיוון שהיא משפיעה ישירות על הביצועים, הסקלביליות והיכולת להכליל למצבים חדשים. הגורמים העיקריים המשפיעים על החלטה זו - ובכלל זה, האם לפתח מודל ייעודי לארגון או לבצע התאמות במודל שאומן מראש - הם אלה:

- זמינות נתונים: כמות הנתונים, איכותם והרלוונטיות שלהם.
- מומחיות טכנית: מיומנויות הארגון בלמידת מכונה ובתחום העיסוק (domain experts).
- משאבי מחשוב: היכולת לאמן ולהריץ מודלים.

לא מכניסים הטיות חדשות או פוגמים בביצועי המודל. איכותם של מאגרי נתונים סינתטיים תלויה במידה שבה הם משקפים את מקרי השימוש המיועדים ואת גיוון התרחישים בעולם האמיתי.

4.2.2.1 סיכונים שיש להביא בחשבון

- נתונים מורעלים הנכללים במודלים שפותחו מחוץ לארגון, בבסיס הידע של מערכת RAG או במועד השאילתה.
- הערכה בלתי אמינה של איכות תוצרי המודל.
- הגברת הטיות או בעיות ביצוע אחרות כאשר משתמשים בנתוני בדיקה סינתטיים שנוצרו על ידי LLM.

4.2.2.2 בקורות מצופות

- שימוש במודלים חיצוניים ממקורות מהימנים, עם תיאור של נתוני האימון היכן שזה אפשרי.
- אימות איכות הנתונים בנתוני קלט שנאספו אוטומטית.
- אימות נתונים סינתטיים כדי למנוע חיזוק הטיות קיימות או הכנסת הטיות חדשות, או פגיעה בביצועי המודל.
- אימות אנשי של נתונים סינתטיים שנוצרו על ידי LLM, אם הדבר ישים.

4.2.3 סוגיות של פרטיות בהקשר של מערכות בינה מלאכותית

השימוש בנתונים אישיים בבינה מלאכותית כפוף לדרישות קפדניות המעוגנות בחוק, לרבות הגבלות על המטרה, מזעור נתונים, מידתיות ושקיפות. בסעיף 3.3.3 אנו דנים בדרישות החלות בתחומי שיפוט ספציפיים.

האחריות לציית לחוק מוטלת על הארגון המפתח את מערכת הבינה המלאכותית או על הארגון שמטמיע אותה (המשתמש). המבקרים נדרשים לבחון תיעוד כגון הערכת השפעות על הגנת המידע (DPIA), כדי לוודא שאכן מקיימים את חובות הגנת המידע. יש להקדיש תשומת לב לכל סוגי הנתונים ששימשו במהלך הפיתוח, לא רק לתוצרי המודל הסופי. לדוגמה, מאפיינים (features) שנבדקו אך בסופו של דבר לא הוכנסו לשימוש עדיין יכולים לכלול עיבוד של נתונים אישיים.

במקום שמשתמשים בו בנתונים אישיים, יש להעריך את חשיבותם לביצועי המודל. כאשר מבקרים וארגונים קובעים את מידת תרומתם של הנתונים האישיים לרמת הדיוק או לאפקטיביות של מערכת הבינה המלאכותית, הם יכולים להעריך אם השימוש באותם נתונים הכרחי ומידתי ביחס למטרה המיועדת. במקרה שמתעורר ספק או חשד לא-ציות להוראות החוק, יש להפנות את המקרה לרשויות

- אילוצי זמן: ל"ז הפרויקט ודרישות המסירה.
- צורכי הסברתיות: דרישות לגבי שקיפות והיכולת לפרש את תוצאות המודל (interpretability).
- תקציב: משאבים כספיים לצורכי הפיתוח והתפעול השוטף.

לדוגמה, ארגונים עשויים להשוות בין הביצועים של אלה: (1) שיטות פשוטות יותר של למידת מכונה; (2) למידה עמוקה שתעבור תהליך כוונן עדין או מודלי שפה קטנים; (3) LLM הנגיש דרך API. בחלק מן המקרים, שיטות פשוטות יותר של למידת מכונה או מודלים קטנים יותר שעוברים כוונן עדין עשויים לספק ביצועים טובים יותר ממודלים מורכבים או יקרים המיועדים למטרות כלליות. על המבקרים לצפות מהארגון לתעד את הנימוקים לבחירת המודל, ולוודא שבחירתו נאותה מן הבחינה הטכנית, וכי היא מוצדקת ושקופה.

הנדסת מאפיינים - התהליך של הפיכת נתונים גולמיים למשתנים בעלי משמעות עבור המודל - חשובה במיוחד. השיטות ששימשו ליצירה, לחילוף ולבחירה של מאפיינים צריכות להיות מתועדות כהלכה. על המבקרים לשים לב לאופן שבו מאפיינים ספציפיים משפיעים על ביצועי המודל, והאם יש מאפיינים העלולים להכניס מורכבות מיותרת או סיכוני הטיה.

במקרה של בינה מלאכותית יוצרת ומודלי יסוד, התהליכים של עיצוב פרומפטים וניהול גרסאות הם קריטיים, מכיוון שההוראות או ההקשר שנותנים למערכת יכולים להשפיע על התוצרים באופן משמעותי. נוסף על הנדסת פרומפטים, הגדרת הפרמטרים לדגימה ממלאת תפקיד מרכזי בעיצוב התנהגות המודל ובעקבות תשובותיו⁶⁶. תיעוד הגדרות אלה חיוני להדירות, לשקיפות ולאחריותיות.

הערכת גישות של למידת מכונה מונחית מתבצעת באופן טיפוסי על ידי מדידת ביצועים על נתוני הבדיקה. במקרה של מערכות בינה מלאכותית יוצרת ומודלים למטרות כלליות, יש צורך להעריך מדדים אחרים. אחד האתגרים העיקריים בפיתוח ביקורת עבור מערכות אלה הוא הסיכון של הזיות (hallucinations). הזיות מתרחשות כאשר המודל מפיק תוצרים שנשמעים סבירים אך הם שגויים עובדתית, מטעים או חסרי פשר. סיכון זה גבוה במיוחד ב-LLM, שיכול לייצר תשובות נחרצות שאינן מעוגנות בנתונים אמנים. חיוני לאמת את התוצרים כנגד מקורות מהימנים כדי לגלות טעויות ולמנוע מידע שגוי - הדבר יסייע לשמר תחושת ביטחון במערכת. עם המדדים האפשריים להערכת מערכות בינה מלאכותית מודרניות נמנים אלה:

- שיעור ההזיות ורמת הדיוק העובדתית.
- עמידות לפרומפטים או קלטים שונים.
- מדדים של בטיחות ורעילות.
- התאמה לתחומי ידע ספציפיים.
- יעילות הטיפול בהקשר ובנתונים.

אין מדד יחיד שיכול לשקף את איכותו הכוללת של מודל, וההערכה דורשת על פי רוב לאזן בין תחומי עדיפות מתחרים, כמו דיוק לעומת הוגנות. על המבקרים לצפות מהארגון להגדיר ולנמק את מדדי הביצוע שבחר.

הדירות (reproducibility) - היכולת לשחזר תוצאות כשמשתמשים באותם נתונים, פרמטרים וסביבה - היא עדיין אבן יסוד בפיתוח מערכת בינה מלאכותית אמנה. מבקרים מתחילים בדרך כלל בסקירת תיעוד, כדי לבסס קו התחלתי לביקורת. אם התיעוד אינו מספק, אינו ברור או מכליל סתירות, ייתכן צורך בשחזור המודל ו/או הניבויים שלו. כאשר שחזור מלא אינו אפשרי (כגון במודלים גדולים או קנייניים), חשוב לבחון תיעוד מעמיק של פרומפטים, הקשרים לאחזור והגדרות תצורה כדי לתמוך בביקורת.

4.3.1.1 סיכונים שיש להביא בחשבון

- טעויות בהכנת הנתונים או בהנדסת המאפיינים היוצרות הטיות או פוגמות באמינות המודל (ראו גם סעיף 4.2, נתונים).
- הערכת יתר של ביצועי המודל, בפרט אם הבדיקות אינן נאותות או אינן מייצגות.
- חוסר יציבות והעדר עקביות של התוצרים, כולל רגישות לשינויי פרומפטים או לשינויים בהקשר במערכות יוצרות.
- חוסר יכולת לאמת טענות לגבי יכולות של מודל יסוד, במקרים רבים, כאשר לרשות המבקרים עומדים משאבים מוגבלים ואין להם גישה למשקלות המודל.

4.3.1.2 בקורות מצופות

תיעוד מועיל ובקרת גרסאות הם מרכיבים חיוניים לשקיפות, לעקיבות וליכולת לבצע את הביקורת. על המבקרים לצפות למצוא את אלה:

- רשומות ברורות לגבי בחירת המודל, התאמתו לצורכי הארגון והיסטוריית גרסאות.
- הערכות ביצועים המשתמשות במדדי השוואה כלליים ובבדיקות ייעודיות ליישום.

66 פרמטרים כמו "טמפרטורה" או "top-k" שולטים בגיוון ובאקראיות של תשובות המודל. לדוגמה, טמפרטורה גבוהה יותר מגבירה את השונות, ואילו ערכים נמוכים יפיקו תוצאות עקביות יותר.



שיקולים אלה רלוונטיים במיוחד למודלי יסוד ולמודלים יוצרים, שבהם ההבחנה בין נתוני האימון לנתונים התפעוליים לעיתים מטושטשת. מודלים מודרניים לראייה ממוחשבת ומודלי שפה רבים נוצרים על ידי כוונת עדין של מודל יסוד שאומן במקור על מאגר נתונים נפרד, גדול הרבה יותר.

התוצאה היא שהנתונים המשמשים לכוונת עדין הם רק חלק מן המכלול שמעצב את המודל הסופי, מה שמקשה לקבוע במדויק אילו נתונים הם "מחוץ להתפלגות". ראיות בשטח מעלות כי מודלים שהותאמו מתוך מודלי יסוד נוטים לבצע הכללות טובות יותר לנתונים חדשים, לעומת אלה שאומנו מאפס באמצעות למידה מונחית. עם זאת הסיבות להכללה משופרת זו עדיין נחקרות.

טכניקות כמו RAG יכולות לשפר מהימנות על ידי הוספת מידע ממקורות חיצוניים לתוצרי המודל, דבר המסייע להתמודד עם בעיות כמו הזיות או ידע לא עדכני. פיתוחים בארכיטקטורת מודלים, כמו חלונות הקשר מורחבים המאפשרים למודל לשקול כמות רבה יותר של מידע סביבתי כשהוא יוצר תשובות, תומכים בתוצרים מהימנים יותר, מדויקים יותר ושהולמים יותר את ההקשר.

על המבקרים לצפות לראות ראיות לכך שהמהימנות והעמידות הובאו בחשבון מהשלב הראשון, בליווי בקורות נאותות לניהול הסיכונים.

4.3.2.1 סיכונים שיש להביא בחשבון

- איכות קוד נמוכה או היעדר בדיקות, המובילים לתוצרים בלתי יציבים או בלתי אמינים.
- פגיעויות הנובעות מקשרי תלות חיצוניים, כמו עדכוני ספקים או שינויים במודל של צד שלישי.
- סיכונים מידע שגוי, כולל הזיות או הנצחת הטיית.
- נתונים שעברו שינויים מינימליים המובילים להתנהגות מודל מוטעית או בלתי אמינה.

4.3.3.2 בקורות מצופות

- עקביות התוצרים: בדיקת עקביות התוצרים על ידי הרצת אותו פרומפט כמה פעמים, במיוחד במודלים יוצרים. הערכת הדירות באמצעות תתי-קבצים של נתוני אימון או נתונים עצמאיים (שיכולים להיות סינתטיים). אם עולה חשד להתאמת יתר או למאפיינים מיותרים, יש לאמן מחדש את המודל עם כחות מאפיינים ולהשוות את התוצאות.
- רגישות לפרטים אישיים: במודלים המשתמשים בנתונים אישיים, יש לבצע אימון מחדש עם כמות מצומצמת של נתונים אישיים או ללא נתונים אישיים כלל, כדי להעריך את שקלול התמורות בין ביצועים לפרטיות.

• ראיות להדירות, כולל נתונים, פרמטרים ופרטים סביבתיים.

• הערכות סיכונים משולבות שבחנו השפעות תפעוליות וחברתיות.

• במקרה של בינה מלאכותית יוצרת ומודלי יסוד, על התייעוד לכלול עיצוב פרומפטים וניהול גרסאות, הגדרות של פרמטרים, ופרטים על APIs חיצוניים או מודלים מאומנים מראש שהארגון משתמש בהם.

• בחירת המודל: תיעוד ברור של הקריטריונים והבדיקות ששימשו להשוואת המודל, או נימוקים לכך שלא בוצעה השוואה.

• בקרת גרסאות: ניהול גרסאות קפדני עבור הנתונים, תוצרי המודל, פרומפטים, תוצרי ניבוי והקוד שמחבר בין תהליכי המערכת (pipeline code). יש לכלול פרטים על הנדסת פרומפטים, ניהול גרסאות והגדרות דגימה.

• רשומות מקיפות: מידע על ארכיטקטורת המודל, נתוני האימון, מגבלות ידועות, מדדי ביצועים, תהליכי פיתוח והחלטות פריסה.

• בינה מלאכותית יוצרת ומודלי יסוד: תיעוד השיטות המשמשות לגילוי הזיות ולצמצום התופעה, ובדיקת המידתיות של שיטות אלה אל מול הסיכונים המזוהים עם השימוש המיועד במערכת.

• הנדסת מאפיינים: תיאור האופן שבו משתנים נוצרים ונבחרים, בדגש על מאפיינים העשויים לפעול כתחליפים עקיפים לתכונות רגישות, וזאת כדי להבטיח עמידה בסטנדרטים של הגנת המידע ולמנוע הטיית.

4.3.2 הבטחת איכות, מהימנות ועמידות בפיתוח

הבטחת איכות (quality assurance) בפיתוח מתייחסת לאיכות הקוד, איכות המודל וגישות מומלצות להבטחת מהימנות ועמידות לפני פריסת המערכת.

• **מהימנות (reliability)** מתייחסת ליכולת של המודל לספק ביצועים עקביים ונכונים בתנאים הצפויים, בפרט בעת טיפול בנתונים הדומים לאלה ששימשו לאימון המודל.

• **עמידות (robustness)** היא היכולת של המודל לשמור על רמת ביצועיו כאשר הוא עומד בפני קלט בלתי צפוי - כמו נתונים בהתפלגות שונה מהנתונים שעליהם הוא אומן; נתונים "מחוץ להתפלגות" (out of distribution - OOD); רעש; או שינויים בסביבה.

רמת דיוק, יש לחזור שנית על בדיקות ההערכה הרלוונטיות המתוארות כאן.

4.4.1 בדיקות ביצועים וקבלה

מערכות בינה מלאכותית צריכות להיבדק בתנאי מציאות טרם פריסתן, לרבות תיקוף טכני של התוצרים ובדיקות משתמשים של זרימות עבודה המשקפות את השימוש הצפוי במערכת בפועל. הבדיקות צריכות להתבסס על מדדים עסקיים וקריטריוני קבלה (acceptance) שהוגדרו בתחילת הפרויקט.

נקודות תורפה פוטנציאליות ופעולות מתקנות יתועדו. תוצאות בדיקות אלה, לצד ראיות לעמידה בקריטריוני הקבלה, יהוו את הבסיס להחלטה אם לתת אור ירוק לפריסה (ראו סעיף 4.5). האמור חל הן על פיתוח מערכות בארגון, הן על מערכות הנרכשות מבחוץ.

4.4.1.1 סיכונים שיש להביא בחשבון

- מערכת הבינה המלאכותית אינה עונה על הדרישות העסקיות.
- הביצועים פחות טובים כאשר משתמשים בנתונים שטרם נראו.

4.4.1.2 בקורות מצופות

- קריטריוני הקבלה ברורים ומוסכמים מראש.
- המערכת נבדקת עם נתונים מייצגים.
- משתמשי קצה מעורבים בבדיקות הקבלה.
- תוצאות הבדיקות וההחלטות מתועדות.

4.4.2 שקיפות והסבריות

שקיפות והסבריות חיוניות לבניית אמון במערכות בינה מלאכותית ולהבטחת אחריותיות. שקיפות משמעה מתן מידע ברור על השימוש במערכת - איך, מתי ומדוע משתמשים בה. הסבריות היא היכולת לתאר כיצד המערכת מגיעה לתוצרים או להחלטות.

הערכת הסבריות עלולה להיות מאתגרת, במיוחד במקרה של מערכות מורכבות או מצד שלישי. מודלים רבים של בינה מלאכותית פועלים כ"קופסאות שחורות", כך שלא קל להבין את ההיגיון הפנימי שלהן. מערכות של בינה מלאכותית יוצרת מתבססות לעיתים קרובות על הסתברות, כך שהן מקשות על שחזור פלטים ספציפיים או על מעקב שיראה כיצד נוצרה תשובה ספציפית. סעיף 4.3 מסביר כיצד ניתן לנתח מתאמים בין מאפיינים וניבויים כדי להסיק מהם על התנהגות המודל.

רמת ההסבריות הנדרשת תלויה במטרת המערכת ובקהל היעד של ההסבר (כמו רגולטור או אדם מהציבור). ההסבריות אמורה להיות מידתית לרמת הסיכון שמייצגת המערכת. חשוב למצוא את האיזון בין המשאבים הנדרשים

- מעקב פרומפטים והקשר: תיעוד וריאציות בפרומפטים, בהקשר וכן בהקשרי RAG כדי לתמוך בשחזור ממוקד, כאשר שכפול מלא של המודל אינו אפשרי.
- ניתוח מתאמים חיצוניים: אם הקוד או המודל אינם נגישים, יש להשתמש בנתונים הזמינים כדי לנתח מתאמים בין מאפיינים וניבויים של המודל, ומכך להסיק על התנהגותו. שיטות אפשריות כוללות LIME או SHAP.
- הדגמת וריאנטים: יש לבקש מהגוף המבוקר להדגים אימון מחדש או זרימות עבודה פשוטות, כדי להבין את התנהגות המודל בתנאים שונים.
- יציבות של תפעול למידת מכונה (machine learning operations - MLOps), כולל בקרת גרסאות, בדיקות יחידה, סקירות קוד, בדיקות אינטגרציה ובדיקות מקצה לקצה.

4.4 הערכת המודל לפני פריסה

לפני פריסה של מערכת בינה מלאכותית, יש לבצע הערכה כדי לוודא שהמערכת עומדת בסטנדרטים שנקבעו בשלב הצעת הפרויקט ובמהלך פיתוח המודל. שלב זה קריטי כדי לוודא שהמערכת מתאימה למטרתה, פועלת כמצופה ועונה על דרישות החוק וכללי האתיקה. ההערכה צריכה להיות מידתית לסיכונים ולהקשר השימוש, ויש לבצעה באופן חוזר במקרה של שינוי במערכת או בסביבתה.

שימוש בבינה מלאכותית ככלי עזר בתהליכי ההערכה ילך וייעשה נפוץ יותר ככל שמערכות בינה מלאכותית יהיו גדולות יותר ומורכבות יותר, כך שבדיקות והערכות ידניות ייעשו מאתגרות יותר. אחת הדוגמאות היא "LLM כשופט" (LLM-as-a-judge): שימוש ב-LLM להערכה ודירוג של תוצרי מערכות בינה מלאכותית.

על המבקרים לוודא כי כל שימוש נוסף בבינה מלאכותית הוערך באופן מספק על ידי הבעלים של הפרויקט.

פרק זה עוסק בנושאים הבאים:

- בדיקות ביצועים וקבלה
- שקיפות והסבריות
- הוגנות ומזעור פגיעה
- אבטחה
- השפעה סביבתית

תהליך ההערכה מיועד לוודא שמערכת הבינה המלאכותית עומדת בסטנדרטים שנקבעו לגבי ממשל (סעיף 4.1) ופיתוח המודל (סעיף 4.3), וכי היא מתאימה בפריסה לסביבת הייצור. סעיף 4.5 עוסק בעלייה לאוויר של המערכת ותוכנית שינויים לאחר שתהליך ההערכה הושלם בהצלחה. סעיף 4.6 עוסק בניטור שוטף ואימון מחדש לאחר פריסה. אם הניטור לאחר פריסה חושף בעיות של הוגנות, אבטחה או

4.4.3 הוגנות ומזעור פגיעה

תהליכי הפיתוח והבדיקה של מערכת בינה מלאכותית צריכים להבטיח שהמערכת תגרום נזק מזערי ותבטיח הוגנות. הוגנות (fairness) פירושה יחס צודק לפרטים ולקבוצות והימנעות מהשפעות שליליות בשל מאפיינים כמו מגדר, מוצא או מיקום. פלט לא הוגן של מערכת עלול לגרום לאפליה, להפסד כספי או לנחיתות חברתית בדרך אחרת.

ישנן דרכים רבות להגדיר הוגנות ולמדוד אותה. הגישה המתאימה תהיה תלויה במטרת המערכת ובהקשר. על המבקרים לצפות מארגונים מבוקרים לתעד ולנמק את בחירתם במדדי הוגנות, ולהראות כיצד מדדים אלה תואמים את דרישות החוק והמדיניות. יש להתייחס הן להוגנות כלפי הפרט, הן להוגנות כלפי קבוצות.

מערכות בינה מלאכותית כמעט לעולם אינן פועלות במנותק ממערכות אחרות. כאשר מתקיימת אינטראקציה בין כמה מודלים, חשוב להעריך כיצד האינטראקציה עלולה לייצר סיכונים או להגביר סיכונים. התמקדות בהערכות של מודל יחיד עלולה לגרום להערכת חסר של הסיכונים הקיימים במערכת הכוללת.

עבור מערכות נרכשות או מודלי יסוד, יש לספק תיעוד המראה כיצד הטיה זוהתה וטופלה. התיעוד יכלול פרטים על מקורות נתוני האימון, שיטות להפחתת הטיה ומדדי הוגנות. התיעוד ובדיקות ההוגנות שמבצעים מפתחים מצד שלישי צריכים להיות רלוונטיים למקרה השימוש של הארגון ולנתונים שלו.

בדיקות הוגנות יכולות לכלול בדיקות של תרחישים חלופיים (counterfactual testing) ושימוש במאגרי נתונים המשמשים כעוגן להשוואה (benchmark datasets). אלה מסייעים לזהות אם המערכת מתייחסת למקרים דומים באופן עקבי, ללא קשר למאפיינים אישיים מוגנים בחוק (כגון מגדר או דת). בדיקת תרחישים חלופיים יכולה לשמש להערכת הוגנות - משנים משתנים רגישים בנתוני הקלט (כמו מגדר או מיקום) ורואים אם התוצרים משתנים באופן לא הולם. קובצי נתונים להשוואה יכולים לשמש להערכת ביצועים והוגנות של מערכת בינה מלאכותית עבור קבוצות שונות.

הוגנות בסיווג שמבצעת הבינה המלאכותית ניתן להעריך על ידי השוואת מדדי ביצוע שונים עבור קבוצות שונות, כדי לזהות הטיה או אפליה פוטנציאליות. על המבקרים להיות מודעים לכך שאף מודל לא יכול לעמוד בכל קריטריוני ההוגנות בבת אחת, במיוחד כאשר שכיחות הקבוצות שונה. לפרטים נוספים על מדדי הוגנות למודלים של סיווג, ראו נספח 2.

הטיה עלולה להיווצר מנתוני אימון בלתי מייצגים או בעלי איכות ירודה, או מהמבנה של המערכת עצמה. מאחר שביצועי המודל משתפרים ככל שכמות הנתונים גדלה (ראו

לבין רמת ההסבר ההכרחית).

כאשר מערכות נרכשות מגורם חיצוני, על הארגון לבקש מידע מהספק על קלטים במעלה הזרם - כמו ארכיטקטורת המודל, נתוני אימון, תהליכי כוונן עדין, מגבלות ידועות ומדדי ביצועים. כדי לסכם מידע זה, משתמשים על פי רוב בכרטיסי מודל (model cards) וגיליונות נתונים. מודלי יסוד המבוססים על קוד פתוח עשויים לספק גישה לקוד, למשקלות המודל ולמדריכים לשימוש אחראי.

4.4.2.1 סיכונים שיש להביא בחשבון

- חוסר יכולת להסביר או לנמק החלטות, המגבילה את יכולתו של הארגון לשפוט את תחזיות המודל ואת אמינותו.
- אי-ציות לתקנות הדורשות הליכים שקופים וביסוס החלטות (דרישות שכיחות במינהל הציבורי).
- מידע בלתי מספק עבור משתמשי הקצה, כך שאינם יכולים לערער על התוצאות.
- הבנה מוגבלת של הנתונים, המודל והניבויים שעלולה להוביל לאי-גילוי של הטיה ואפליה.
- "Fairwashing" - מניפולציה בטכניקות הסבר בדיעבד כדי להצדיק תוצאות בלתי הוגנות.

4.4.2.2 בקורות מצופות

- ניהול תיעוד על מטרת המערכת, המבנה שלה, תוצריה הצפויים ומגבלותיה. התיעוד יכול לכלול כרטיסי נתונים (data cards - מקורות והטיות של קובצי נתונים), כרטיסי מערכת (system cards - ארכיטקטורה, נתוני אימון, מגבלות) וכרטיסי ביקורת (audit cards - דרישות רגולציה וכימות סיכונים).
- הצגת הודעה בפני משתמשים כאשר הם באינטראקציה עם בינה מלאכותית.
- מתן מנגנונים בידי המשתמשים לערער על החלטות, ווידוא קיומם של הסדרים הנתונים למשתמשים אפשרות לקבל פיצוי אם נפגעו.
- הצדקת השימוש במודלים בעלי "קופסה שחורה" במקרה הרלוונטי.
- הבהרת אחריות משפטית, לרבות דרישות למתן גילוי נאות לגבי תוכן סינתטי.
- שימוש בפתרונות כמו סימני מים (גילויים הממוקמים על גבי תוכן שנוצר באמצעות בינה מלאכותית) וזיהוי מקור ואותנטיות (איתור מקור התוכן וההיסטוריה שלו), כדי לסייע בזיהוי חומר שנוצר באמצעות בינה מלאכותית.

- שימוש בבדיקת תרחישים חלופיים (שינוי משתנים רגישים כדי לבדוק אם מתקבלים שינויי תוצר בלתי הולמים) ובקובצי נתונים להשוואה כדי להעריך הוגנות.
- הבניית פיקוח אנושי לתוך תהליך האימות, במיוחד ביישומים הכרוכים בסיכון גבוה או ברגישות מיוחדת.
- ניטור שוטף המבטיח גילוי בעיות של הוגנות בתפעול לאחר העלייה לאוויר וטיפול בבעיות אלה.
- במקרה של בינה מלאכותית יוצרת, בדיקת פרומפטים ואת תשובות המערכת כדי לוודא שאינה לוקה בדעות קדומות או פוגעניות.
- זיהוי סיכונים אפליה צפויים, מעקב אחריהם וטיפול בהם.
- במודלים של צד שלישי או מודלי יסוד, הארגון המבוקר יודא כי בדיקות ההוגנות והתיעוד המסופקים על ידי הגוף המפתח רלוונטיים למקרה השימוש המיועד.

4.4.4 אבטחה

אבטחה מתייחסת להגנה על נתונים, נכסים ופונקציות של המערכת מפני גישה בלתי מורשית, ניצול לרעה או נזק. ככל שמערכות בינה מלאכותית מרחיבות את יכולותיהן ונעשות אוטונומיות יותר, כך עולים סיכונים הבטיחות והאבטחה. הדבר נכון במיוחד במקרה של מערכות שמסוגלות לפעול באופן עצמאי או לעבד מידע רגיש.

מערכות בינה מלאכותית מתמודדות עם אותם אתגרי אבטחה שקיימים במערכות IT אחרות. אולם כמויות הנתונים הגדולות וכוח המחשוב האדיר המעורבים בפיתוח בינה מלאכותית משמעם שהאבטחה של תשתית מבוססת ומבוססת ענן חשובה במיוחד.

4.4.4.1 סיכונים שיש להביא בחשבון

- מערכות בינה מלאכותית יכולות לקבל גישה לנתונים אישיים או סודיים, לחשוף אותם או להשתמש בהם באופן שגוי. קיים גם סיכון של ביצוע שינויים בלתי מורשים, גרימת נזק או אובדן נתונים.
- אי-ציות לדיני הגנת המידע, כמו GDPR, עלול להוביל להשלכות משפטיות ולפגיעה במוניטין. סיכון זה גדל אם הנתונים מעובדים או מאוחסנים במדינות שקיימות בהן תקנות אסדרה שונות.
- גורמים זדוניים עלולים לבצע פעולות מניפולטיביות בקלט הנתונים כדי להטעות את המודל, ולהוביל לפלטים שגויים או מזיקים. דוגמאות לכך כוללות הרעלת נתונים ומתקפת הזרקת פרומפטים.

סעיף 4.2), התוצאות משקפות לא פעם העדפה לקבוצות רוב, בשעה שמיעוטים נותרים בעמדת נחיתות. מודלים מסוג בינה מלאכותית יוצרת פגיעים לכך במיוחד, מפני שהם מאומנים בדרך כלל על קובצי נתונים גדולים ובלתי מסוננים מהאינטרנט שעשויים לשקף הטיות חברתיות. התוכן שנוצר יכול לשקף או לחזק סטריאוטיפים מזיקים.

כאשר המבקרים רוצים להעריך אם קיימת הטיה בנתוני האימון, עליהם לבחון את המקור והאיכות של מאגרי הנתונים, במיוחד אם הם נלקחו מהאינטרנט או נוצרו באופן סינתטי. על הארגון להראות כי הוא טיפל בנושאים של זכויות קניין רוחני, הגנת המידע וסטנדרטים של אתיקה בעת איסוף הנתונים והשימוש בהם. תיעוד ברור צריך להיות זמין למבקרים.

הפיקוח האנושי חשוב, בפרט ביישומים הכרוכים בסיכון גבוה או ברגישות מיוחדת. על המבקרים לוודא כי בודקים אנושיים מעורבים באימות הפלט (אימות מסוג "אדם בתוך התהליך" - human in the loop), במיוחד כאשר השכעתן של החלטות אוטומטיות עשויה להיות משמעותית. על הארגון להראות כי הוא מקיים ניטור שוטף לגילוי בעיות של הוגנות וטיפול בהן מרגע שהן מתעוררות לאחר עליית המערכת לאוויר (ראו מידע נוסף בסעיף 4.6).

4.4.3.1 סיכונים שיש להביא בחשבון

- מערכת הבינה המלאכותית מחזקת או מרחיבה פערים קיימים.
- הטיה או אפליה אינן מתגלות בשל בדיקות הוגנות בלתי מתאימות או בלתי מספקות.
- אינטראקציה בין מערכות או מודלים שונים יוצרת סיכונים חדשים שאינם מתגלים על ידי הערכת כל רכיב בנפרד.
- מערכות שנרכשו מבחוץ או שייכות לצד שלישי נעדרות שקיפות לגבי בדיקות הוגנות או הפחתת הסיכון להטיות.
- הסתמכות על מדד הוגנות יחיד או בסיסי השוואה גנריים עלולה להקנות תחושת ביטחון כוזבת במהימנות התוצרים.

4.4.3.2 בקורות מצופות

הוגנות היא בין השיקולים החשובים ביותר לאורך שלבי התכנון, הפיתוח והפריסה של מערכות בינה מלאכותית. על המבקרים לצפות לראות ראיות לבדיקות הוגנות עמידות ותואמות הקשר, תיעוד ברור וניטור שוטף. כאשר אין אפשרות להבטיח הוגנות באופן מוחלט, חשוב להכיר בסיכונים בשקיפות ולנהל אותם.

- תיעוד של מקורות נתוני האימון, שיטות להפחתת הטיות ומדדי הוגנות. הכללת הרציונל לפי המקרה.

- תוקפים עלולים להצליח לחדור למודלים או לקוד הבסיס, ולבצע בהם מניפולציות או שימוש לרעה.
- מערכות בינה מלאכותית בעלות רמת אוטונומיות גבוהה, במיוחד סוכני AI, עשויות לקבל החלטות או לבצע פעולות ללא פיקוח אנושי מתאים.

4.4.4.2 בקורות מצופות

נושא האבטחה נשקל בכל שלב במחזור החיים של הבינה המלאכותית. הערכות סיכונים וסקירות קבועות מתבצעות כדי לוודא שהבקורות עדיין מועילות ככל שהטכנולוגיה והאיומים מתפתחים. דיון נוסף על ניטור אבטחה שוטף מובא בסעיף 4.6.

מגוון המערכות והיישומים המבוססים על בינה מלאכותית מקשה לחזות מראש את כל הסיכונים הפוטנציאליים. ארגונים נדרשים לבצע הערכת סיכונים יסודית ולהיות מודעים לסיכוני האבטחה הספציפיים העשויים להתעורר בשלבים שונים של מחזור החיים של המערכת - מהפיתוח ועד הפריסה. האבטחה צריכה להיות שיקול כאשר מחליטים אם לפתח מערכת בינה מלאכותית בארגון או לרכוש מצד שלישי. במקרה של סוגיות אבטחה ידועות, יש ליישם אסטרטגיות להפחתת הסיכונים ולבדוק את יעילותן. כל מערכת צריכה לעמוד לסקירות קבועות כחלק מאסטרטגיית הפחתת הסיכונים, בפרט במקרה של התפתחויות חדשות.

בקורות אפשריות:

- מניעת גישה בלתי מורשית לנתונים, למודלים ולקוד באמצעות בקורות גישה, ניטור רשתות וקריפטוגרפיה (כולל הצפנה).
- הפרדת סביבת פיתוח הבינה המלאכותית מרשתות ה-IT הראשיות, כדי להקטין את הסיכון של פגיעה נרחבת.
- סקירות קבועות של המערכות לגילוי פרצות וסימני תקיפה. בדיקת בעיות בפגיעות המערכת ועדכון אמצעי ההגנה לפי הצורך (צוותים אדומים).
- הכשרת אנשי צוות הארגון לזהות איומי אבטחה ולהבין את אחריותם.
- יישום אמצעי הגנת מידע כדי להבטיח ציות לחוקים ולתקנות הרלוונטיים. לדוגמה, בדיקות שמסננות נתונים זדוניים טרם כניסתם למערכת, ובדיקות של נתונים יוצאים טרם יציאתם מהמערכת (בדיקות אלה עונות על צורכי הבטחת איכות והגנת מידע). במקרה של בינה מלאכותית יוצרת, שימוש במסנני תוכן למניעת חשיפה של מידע רגיש.

- בדיקת עמידתם של ספקי שירותי ענן וספקי מודלים חיצוניים בתקני האבטחה, ובדיקת הסכמי רמת השירות כדי לוודא שהם מכסים ניהול תקריות ותמיכה.
- קביעת נהלים לגילוי תקריות אבטחה, דיווח עליהן ותגובה, כולל חזרה למצב קודם (rollback) או הוצאה משירות במקרה הנדרש.

4.4.5 השפעה סביבתית

השפעתן הסביבתית של מערכות בינה מלאכותית, במיוחד LLM, יכולה להיות משמעותית. האימון והפריסה של מערכות אלה מצריכים על פי רוב כוח מחשוב מהותי, המוביל לצריכת אנרגיה גבוהה ולפליטות פחמן מוגברות. ככל שקהלים רחבים יותר מאמצים בינה מלאכותית, כך גוברת החשיבות של הבנת טביעת הרגל הסביבתית של מערכות מסוג זה וניהולה.

כאשר מערכת נרכשת מצד שלישי, על המבקרים להתייחס לביצועים הסביבתיים של הספק - בין היתר שימוש באנרגיה מתחדשת, תשתית חסכונית באנרגיה ופרקטיקות עסקיות המבוססות על תפיסת הקיימות. רבים מהספקים של שירותי ענן כיום מציעים כלים להערכת טביעת הרגל הפחמנית של פעילויות מחשוב ענן.

4.4.5.1 סיכונים שיש להביא בחשבון

- עלות סביבתית שעולה על יתרונותיה של מערכת הבינה המלאכותית.
- אי-ציות לתקנים של דיווח סביבתי או לחקיקה.

4.4.5.2 בקורות מצופות

על המבקרים לחפש תיעוד המוכיח ציות לתקני דיווח סביבתי או לחקיקה אחרת. מאחר שהמתודולוגיות למדידת השפעה סביבתית עדיין אינן מפותחות במידה מספקת, ביקורת חוקרת עלולה לעורר אתגרים. ככל שהתקנים והכלים למדידת השפעה סביבתית ימשיכו להתפתח, על המבקרים להתעדכן בגישות המומלצות ובדרישות החדשות בתחום זה.

- תיעוד של הציות לתקני דיווח סביבתי ולחקיקה הרלוונטית.
- בחינת הביצועים הסביבתיים של ספקים מצד שלישי.
- שימוש בכלים זמינים להערכת טביעת הרגל הפחמנית של מערכות בינה מלאכותית והמידה שבה הן מנצלות משאבים.
- הערכה ותיעוד של צריכת אנרגיה, צריכת מים ופליטות גזי חממה, היכן שניתן לעשות זאת.

4.5 פריסה וניהול שינויים

פריסת מערכת בינה מלאכותית היא שלב קריטי שמעביר את הטכנולוגיה משלב הפיתוח אל העלייה לאוויר. שלב זה מצריך תכנון קפדני כדי להבטיח שהמערכת תשתלב בצורה חלקה בתהליכים העסקיים ותספק את התועלת המצופה ממנה. ניהול שינויים אפקטיבי מסייע לארגונים להסתגל לדרכי עבודה חדשות ולצמצם את הסיכון להפרעות. פריסה מוצלחת צריכה גם לאפשר עיצוב מחדש של תהליכים ותמיכה במימוש היתרונות המצופים.

פרק זה עוסק בנושאים הבאים:

- מוכנות עסקית וניהול שינויים.
- קריטריונים לעלייה לאוויר ושערי איכות.

פריסה וניהול שינויים קשורים בקשר הדוק לשלבים אחרים במחזור החיים של בינה מלאכותית. הנושאים של הצדקה עסקית, מטרות הפריקט ומעורבות בעלי עניין הנדרשת לפריסה מפורטים בסעיף 4.1. סעיפים 4.3 ו-4.4 עוסקים בתכנון טכני, בדיקות משתמשים וקריטריוני קבלה שעליהם מבוססת הפריסה. סעיף 4.6 מתייחס לניטור שוטף, אימון מחדש ומעקב ציות לאחר הפריסה.

4.5.1 מוכנות עסקית וניהול שינויים

פריסת מערכות בינה מלאכותית דורשת לעיתים קרובות שינויים משמעותיים בתהליכי העבודה ובמבנה הארגוני. כדי לממש את כל יתרונותיה של הבינה המלאכותית, ארגונים נדרשים לעיתים לעצב מחדש תהליכים קיימים ולוודא שזרימות העבודה החדשות תואמות את יכולות המערכת החדשה. תהליך זה יכול לכלול מיפוי של התהליכים הנוכחיים וזיהוי הזדמנויות לאוטומציה או שיפור. יש להגדיר בבירור את התפקידים האנושיים ואת אלה של הבינה המלאכותית, ולטפל בצעדים מיותרים או צווארי בקבוק באופן מיטבי.

כל בעלי העניין צריכים לקבל מידע על המטרה והפונקציה של המערכת ולהבין אותן. עם בעלי העניין נמנים כל משתמשי הקצה, צוותי IT וההנהלה. הדרכת העובדים היא היבט חיוני במוכנות הארגון, ובמיוחד הדרכה לגבי אי-הוודאות במערכות הסתברותיות, "הזיות" של LLM ואמצעי הבטחת איכות נאותים. אסטרטגיות תקשורת אפקטיביות חשובות לטיפול בעלות ואמון בקרב בעלי העניין. מנגנונים כמו התייעצויות עם הציבור וסדנאות משותפות יכולים לשמש ככלים לעידוד המעורבות של בעלי עניין.

מסמכי מדיניות בנושא בינה מלאכותית, כמו אסטרטגיית בינה מלאכותית פנימית או קווים מנחים לשימוש בטיחותי ואתי, יכולים לסייע להכין עובדים בכל חלקי הארגון להסתגל לפתרונות מבוססי בינה מלאכותית. מסמכים אלה אמורים להבהיר את התפקידים ותחומי האחריות של כל אחד מהעובדים ולוודא שמשתמשי הקצה מודעים למגבלות

המערכת. צריכות להיות הנחיות המבהירות את יחסי הגומלין בין מערכות בינה מלאכותית לעובדים אנושיים, כולל סמכות ואחריותיות במקרה של מחלוקת.

יש חשיבות קריטית להגדרת תהליכי המעבר מהפיתוח לייצור, וכן לתכנון המשכיות והסדרי מסירה למועד שבו אנשי מפתח יעזבו את הפרויקט. על הארגון לשקול להקים פורום של בעלי המומחיות הטכנית הנדרשת - נפרד מצוות הפיתוח - כדי להעריך באופן ביקורתי את השימוש בתוצרי המודל: הן יחידות של בקרה פנימית, הן מנגנונים לטיפול בתלונות של משתמשים או בסוגיות הקשורות לנתונים.

כדי להבטיח שיתרונות המערכת יתממשו, על הארגון לקבוע יעדים ברורים (כמו רווחי התייעלות, חיסכון בעלויות, איכות שירות משופרת או שיעורי טעות מופחתים) ולפתח אינדיקטורים והסדרי ניטור שיעקבו אחר מידת התממשות היתרונות המצופים לאחר הפריסה. האחריות למימוש היתרונות צריכה להיות מוטלת על גורם ספציפי, ויש להפיק לקחים וללמוד מהם כבסיס לפרויקטים עתידיים.

4.5.1.1 סיכונים שיש להביא בחשבון

- ריכוז הידע הטכני על מודל מסוים אצל מספר קטן של עובדים או יועצים חיצוניים, דבר המוביל לחוסר תקשורת וחוסר יעילות בהטמעת המערכת.
- משתמשים במערכת החדשה שאינם מוכנים לשנות תהליכי עבודה מוכרים או מרגישים עודף ביטחון או סקפטיות לגבי היכולות של מערכת הבינה המלאכותית.
- תהליכי עבודה נוקשים או חסרי גמישות המעכבים את אימוץ המערכת או יוצרים חסמים לכך.
- היעדר שקיפות והסברתיות עבור משתמשי הקצה.
- מעבר לא תקין מהפיתוח לייצור.
- משתמשי קצה שאינם מודעים במידה מספקת למגבלות מערכת הבינה המלאכותית, דבר המוביל לחוסר שקיפות או חוסר הוגנות בהחלטות.
- הדרכה או הנחיות בלתי מספקות למשתמשי הקצה, המובילות לשימוש שגוי או חוסר הבנה.
- היעדר תהליכים מוגדרים למעבר, תכנון המשכיות ומסירה בסיום תפקיד.
- תקשורת לא מסודרת ומעורבות לקויה של בעלי עניין, המובילות להתנגדות או לחוסר אמון.
- היעדר תכנון מחדש של התהליכים, או היעדר הגדרת יעדים למימוש יתרונות המערכת. התוצאה היא החמצת הזדמנויות לשיפור.

4.5.1.2 בקרות מצופות

- מתן הדרכות למשתמשים כדי להבטיח שיהיו להם המיומנויות וההבנה הנדרשות כדי להשתמש במערכת בצורה מועילה, כולל כישורי הערכה ביקורתית.
- גיבוש מדיניות או הנחיות ברורות בנוגע לאינטראקציה בין האדם לבינה המלאכותית, המבהירות נושאים של סמכות ואחריות.
- הגדרת תהליכים עבור המעבר משלב הפיתוח לייצור, תכנון המשכיות ומסירת המידע הלאה בסיום תפקיד.
- הקמת פורומים לבקרה פנימית בלתי תלויה וטיפול בתלונות חיצוניות.
- שימוש באסטרטגיות תקשורת כדי לעורר מעורבות בקרב בעלי עניין ולבנות אמון.
- מיפוי ועיצוב מחדש של תהליכים עסקיים כדי לשלב את הבינה המלאכותית בצורה מועילה.
- קביעת יעדים למימוש יתרונות המערכת ומעקב התקדמות.
- הקצאת אחריות למימוש יתרונות ושיפורים מתמידים.

4.5.2 קריטריונים לעלייה לאוויר ושערים

עלייה לאוויר תלויה בכך שהמערכת עברה את כל ההערכות טרום פריסה, כולל עמידה בקריטריוני קבלה בהיבטים הטכניים, האתיים והמשפטיים (ראו סעיפים 4.3 ו-4.4). תוצאותיהן של הערכות אלה, לרבות בדיקות קבלה על ידי המשתמשים וראיות לעמידה במדדי KPI עסקיים, חייבות להיות מתועדות ומאושרות בחתימת בעלי עניין רלוונטיים.

הקריטריונים לעלייה לאוויר צריכים לכלול אישור לכך שהמערכת מצטיינת לחוקים ולכללים הנדרשים, חבילות תיעוד (כגון כרטיסי מודל, כרטיסי מערכת וכרטיסי ביקורת) ופרסום מידע רלוונטי לגבי שקיפות ואחריות (ראו סעיף 4.4). חייבים להיות תהליכי ניהול שינויים הקובעים כיצד ינוהלו עדכונים עתידיים, הגירת נתונים וחזרה למצב קודם - כולל בקרת גרסאות, בדיקות תאימות ומנגנונים של חזרה למצב קודם המבטיחים המשכיות ומהימנות.

כאשר מערכות בינה מלאכותית נרכשות מבחוץ או מסתמכות על מודלים חיצוניים, הסכמי רמת השירות צריכים להבטיח תמיכה שוטפת ושליטה על תשתיות או שירותים חיצוניים במידה מספקת לניהול תקריות וניתוח כשלים (ראו סעיף 4.6).

4.5.2.1 סיוכונים שיש להביא בחשבון

- מערכת שאינה בשלה או לא עמדה ברף הביצועים הנדרש, ולא עמדה בקריטריונים להעלאתה לאוויר - ובכל זאת עוברת לסביבת הייצור.
- היעדר אפשרות לשחזר התנהגות של גרסאות ייצור קודמות לאחר שגרסה חדשה עולה לאוויר.
- תמיכה בלתי מספקת או שליטה בלתי מספקת במערכות שנרכשו או פותחו מחוץ לארגון.

4.5.2.2 בקרות מצופות

- וידוא קיומם של הקריטריונים לעלייה לאוויר ושערים (נקודות בקרה), כולל אישורי ציית וחבילות תיעוד.
- חתימה על הסכמי רמת שירות עבור מערכות שנרכשו או פותחו מחוץ לארגון, המבטיחים תמיכה ושליטה באופן רציף.
- ניהול בקרת גרסאות, בדיקות תאימות ומנגנוני חזרה למצב קודם לניהול שינויים (ראו סעיף 4.6).

4.6 מערכות בינה מלאכותית בסביבה תפעולית

מרגע שמערכת בינה מלאכותית נפרסת בארגון, היא פועלת בתנאי העולם האמיתי, אשר עשויים להיות שונים מאלה ששררו במהלך שלבי הפיתוח והבדיקה. היות שכך, מערכות בסביבה התפעולית חשופות לכמה סיכונים חדשים.

נתונים "חיים" לעיתים שונים ממאגרי נתונים היסטוריים או כאלה שנאספו ממקומות שונים, ששימשו לאימון ולבדיקה. נתונים חיים יכולים לכלול קלט בלתי צפוי או "מלוכלך". מרגע שבינה מלאכותית הופכת לחלק מהתהליכים התפעוליים של הארגון, משתמש בה צוות שלא בהכרח היה מעורב בפיתוחה. הדבר מגביר את הסיכון לשימוש שגוי או שימוש למטרות החורגות מהכוונה המקורית. סביבות תפעוליות גם כרוכות בדרישות גדולות יותר מבחינת היציבות הטכנית, המהירות והזמינות.

עם הזמן, שינויים בהתנהגות המשתמשים, גורמים דמוגרפיים או גורמים חיצוניים יכולים לשנות את התפלגות נתוני הקלט. מצב זה ידוע כ"סחף נתונים" (data drift). גם סחף של המודל יכול להתרחש - מצב שבו ביצועי המודל הולכים ונגרעים לאורך זמן. שינויים אלה עלולים לפגוע בביצועי המערכת או להכניס בה הטיות חדשות. כמו כן שינויים בנורמות המקובלות, במדיניות או בחקיקה עשויים להשפיע על דרישות הציית.

אפקטיביות הניטור והניהול חיונית כדי להבטיח שהמערכת תמשיך לספק ערך, תמשיך לציית לתקנות ותפעל בצורה בטוחה ואיתנה. ארגונים נדרשים לבחון תצפיות בקביעות ולדרוש מהבעלים של המערכת בארגון להיות מודעים

- שינויים מטעם הספק ושינויי גרסה של המודל: ספקים עשויים לעדכן מודלים או להוציא אותם מהשוק, לשנות APIs או לשנות את תנאי השירות. אירועים אלה יכולים לפגום בביצועים או לעורר סיכונים חדשים.

4.6.2 בקורות מצופות

כדי לנהל סיכונים אלה, על הארגון לבצע את הפעולות האלה:

- מעקב רציף אחר רמת הדיוק, האמינות, העמידות ומדדי הביצוע העיקריים שנקבעו בשלב תכנון הכריזקט.
- מעקב אחר איכות נתוני הקלט וייצוגיות.
- קביעת בדיקות אוטומטיות או מתוזמנות שתכליתן לבדוק אם מתרחש סחף נתונים או סחף מודל, והגדרת רמה קבילה של סיפי ביצוע. אם מערכת הבינה המלאכותית פותחה על ידי צד שלישי ואין בידי הארגון פרטים על הנתונים, עליו לקבל מהצד השלישי דוחות קבועים על איכות הנתונים.
- גיבוש מדיניות לשימור או עדכון של מודלים במקרה שהתגלה סחף. תיעוד אירועי אימון מחדש, כולל הנתונים ששימשו לאימון, השינויים שנעשו בהם ותוצאות ההערכה. וידוא כי שינויים במאפיינים, בפרומפטים או בארכיטקטורת המודל מסומנים במספרי גרסאות ומתועדים כראוי.
- מעקב אחר הטיות והוגנות בפלטים של המערכת, לצד מעקב אחר משוב ותלונות משתמשים.
- קביעת נהלים לגילוי, רישום וחקירה של תקריות. הודעה לבעלי העניין לפי הנדרש ונקיטת פעולה מתקנת, כולל החזרת המערכת למצב קודם או הוצאתה משירות לפי הצורך.
- סקירות קבועות של ציות המערכת לחוקים ולתקנות עדכניים, במיוחד עבור מערכות שמעבדות נתונים אישיים, מקבלות החלטות המשפיעות על פרטים או מופעלות במגזרים מפוקחים.
- מתן הנחיות ברורות והדרכה למשתמשים, כדי לוודא שהם מבינים מהו השימוש המיועד של המערכת ומהן מגבלותיה. יש לזהות שימושים שגויים שניתן לצפותם מראש באופן סביר, לתעד אותם ולנהל אחריהם מעקב.
- הבטחת בעלות ברורה על הניהול השוטף, ניטור ועדכונים.
- הגדרת טריגרים להוצאה משירות, נהלים להעברת נתונים לארכיון או מחיקתם, ותחומי אחריות לניהול מעבר חלק ליום שבו המערכת תוצא משימוש.

לביצועיה. אם מערכת הבינה המלאכותית אינה עומדת עוד בדרישות הביצוע או הציות, או שהצורך העסקי השתנה, יש לנקוט פעולה מתקנת (כגון אימון מחדש, התאמות טכניות או הוצאת המערכת משירות). אימון מחדש צריך להתבצע בהתאם לאותם סטנדרטים של תיעוד, בדיקה וקבלה שחלו בעת פיתוח המודל הראשוני (ראו סעיף 4.3). שינויים של מאפיינים, פרומפטים או ארכיטקטורת המודל חייבים להיות מסומנים במספרי גרסאות ומתועדים כראוי.

מערכות בינה מלאכותית רבות מסתמכות על מודלים, APIs או שירותי ענן מצד שלישי. ספקים עשויים לעדכן מודלים או להוציא אותם מהשוק, לשנות APIs או לשנות את תנאי השירות. שינויים אלה עלולים להשפיע על הביצועים או להכניס סיכונים חדשים. אם עדכון על ידי הספק או שינוי המודל משפיעים על התנהגות המערכת באופן מהותי, על הארגון להפעיל שוב את קריטריוני הקבלה, ולהשיב את המערכת לסביבה התפעולית רק לאחר שהיא עוברת את המבחנים הרלוונטיים.

4.6.1 סיכונים שיש להביא בחשבון

- סחף ביצועים (performance drift): המערכת עשויה להתקל בנתונים חיים שהם שונים מנתוני האימון או הבדיקה שסוננו בקפידה, מה שיוביל לירידה ברמת הדיוק או האמינות של המערכת. ייתכן שהבדלים אלה יהיו נוכחים מההתחלה, וייתכן שיופיעו בהדרגה בצורת סחף נתונים.
- שימוש שגוי או לא בהתאם לכוונה: מערכות בינה מלאכותית יכולות לשמש בדרכים שאינן צפויות מראש במהלך הפיתוח. משתמשים חדשים עשויים ליישם את מערכת הבינה המלאכותית בהקשרים שונים או לטעות בפרשנות התוצרים שלה. זה עלול להוביל להחלטות לא נאותות ולשימוש בניבויים של המערכת מחוץ להקשר שנועדה לו.
- ירידת ביצועים עקב שימוש שגוי: מערכת הבינה המלאכותית עשויה לספק ביצועים כמצופה, אך להוסיף להיכשל בהשגת ה-KPIs אם ההליכים סביבה אינם מיושמים כראוי. לדוגמה, ייתכן שהיא לא תספק את התועלת המצופה ממנה אם העובדים לא משתמשים בניבויים שלה בצורה מועילה.
- היעדר בעלות: ללא אחריות ברורה על הניהול השוטף, כולל בעלות עסקית והעברת אחריות כשאדם עובר תפקיד, ייתכן שהמערכת לא תנטר כראוי או לא תעודכן במועד הנדרש.
- פערי ציות: חוקים ותקנות עשויים להשתנות עם הזמן, ומערכות יכולות להפוך לכאלה שאינן עומדות בחוק אם לא מקיפים על סקירות יזומות.

נסכחים

נספח 1: רכיבים של ביקורת טכנולוגיית מידע קלאסית בהקשר של בינה מלאכותית

מערכות בינה מלאכותית חולקות מאפיינים רבים עם תוכנה מסורתית, כך שתקני ביקורת מבוססים כמו התהליך התקני חוצה התעשיות לכריית מידע (Cross-Industry Standard Process for Data Mining) יכולים להנחות את סקירתן⁶⁸. התקן משקף תהליך פיתוח סטנדרטי של מערכות בינה מלאכותית, גם אם הגורם המפתח לא משתמש בו באופן מפורש. מבקרי IT יכולים אפוא לבצע סקירה ברמה גבוהה ללא מומחיות בבינה מלאכותית. התקן מחולק לשישה שלבים:

1. הבנה עסקית
2. הבנת הנתונים
3. הכנת הנתונים
4. בניית המודל ופיתוח
5. הערכת המודל לפני פריסה
6. פריסה ותהליכי ניהול השינויים המלווים אותה

הביקורת בחלק מן השלבים, כמו בשלב ההבנה העסקית, מתבצעת באותו האופן כמו ביקורת של פרויקטים רגילים לפיתוח תוכנה. שלבים אחרים, כמו הכנת הנתונים ובניית המודל, עשויים לחייב מומחיות בנושא. הגישה המומלצת היא להרכיב צוותי ביקורת המכילים תערובת של מיומנויות. צוות מאוזן יכול לכלול מבקרים מומחים בתחום, מבקרי IT ומדעני נתונים. אם חסר בצוות ניסיון בתחום מדעי הנתונים, על אנשי הצוות לבקש חוות דעת מעמיתיהם שהם בעלי מומחיות זו. באותו אופן, צוותים שהיכרותם עם ביקורות IT טובה פחות, צריכים לערב מבקרי IT מנוסים. גישה זו תסייע להבטיח שכל היבטי המערכת ייסקרו כראוי.

בית הדין הפדרלי לביקורת חשבונות של גרמניה (Bundesrechnungshof) נקט גישה זו בביקורת הראשונה שביצע למערכת בינה מלאכותית. צוות הביקורת כלל מבקר מומחה בעל ידע על הארגון המבוקר ושני מבקרים טכניים: אחד בעל רקע ב-IT, והאחר - במדעי הטבע. מבקרים אלה הובאו מיחידות אחרות כדי להשתתף בצוות ייעודי לביקורת בינה מלאכותית.

כלי העזר לביקורת הכלול במסמך זה נבנה בהתאם למחזור החיים של הבינה המלאכותית, ומציג שאלות ביקורת המתאימות למבקרים מומחים, מבקרי IT ומדעני נתונים.

- מעקב אחר גרסאות הספק/מודל שבשימוש. בדיקות שמטרתן גלולת נסיגה בביצועים או בעיות תאימות לאחר עדכוני הספק. ניהול מנגנונים המאפשרים להחזיר את המערכת למצב קודם או פתרונות חלופיים, ותיעוד כל השינויים והשפעתם.

5. סיכום ומסקנות

גופים במגזר הציבורי מפתחים ומטמיעים מערכות בינה מלאכותית בשירותים לציבור כדי לשפר את איכות השירות ולהפחית עלויות. ואולם טכנולוגיה חדשה זו מלווה באתגרים חדשים. מוסדות ביקורת עליונים נדרשים להצטייד בכל הדרוש כדי לבצע ביקורת מועילה במערכות בינה מלאכותית. כמה מדינות כבר החלו להפעיל תוכניות פיילוט או שכבר ביצעו את הביקורות הראשונות במערכות מסוג זה במגזר הציבורי. מסמך זה נועד לסייע לקהילת הביקורת הבין-לאומית על ידי התוויית הנחיות ושיטות מומלצות לביצוע ביקורת במערכות בינה מלאכותית.

קטלוג הביקורת בנוי סביב שישה תחומי מפתח בביקורת, המתמקדים בניהול וממשל של הפרויקט, נתונים, פיתוח המערכת, הערכה לפני פריסה, פריסה וניהול שינויים, וניהול שוטף של מערכות בינה מלאכותית בסביבה תפעולית. הקטלוג מנחה את המבקרים לאורך תהליך טיפוס של פיתוח בינה מלאכותית. טכניקות הביקורת המתוארות כאן יושמו בביקורות שביצעו המוסדות החתומים על מזכר ההבנות. מטרת הביקורות הייתה להעריך את התועלת והמהימנות של יישומי למידת מכונה, וכן את היעילות והאפקטיביות של ההטמעה והתפעול של יישומים אלה.

בינה מלאכותית נמצאת בשימוש במגוון רחב של מגזרים, ובכל אחד מהם סיכונים שונים החלים על יישומים שונים. תחום הבינה המלאכותית עודנו מתפתח, חקיקה חדשה ממשיכה להתגבש ומוקמים גופי אסדרה ופיקוח ייעודיים. קטלוג ביקורת זה משקף את התקן לביקורת בינה מלאכותית בקרב מוסדות הביקורת העליונים שחיברו את המסמך, במועד הכתיבה⁶⁹. ייתכן שהמסמך יעודכן ככל שיצטבר ניסיון ביקורת נוסף ויתקבלו תוצאות של מחקרים חדשים.

למסמך זה מצורף כלי עזר לביקורת בינה מלאכותית המאפשר למבקרים לבחור שאלות מתוך רשימה מוכנה וליצור שאלון מותאם ההולם את הביקורת המסוימת שהם מבצעים. כלי העזר לביקורת מספק שאלות ביקורת מומלצות, ומציין מה הן ראיות הביקורת הנדרשות.

67 עדכון אחרון: דצמבר 2025.

68 A. Clark (2018): [The Machine Learning Audit-CRISP-DM Framework](#) 68

נספח 2: נתונים אישיים והתקנות הכלליות להגנת המידע (GDPR) בהקשר של בינה מלאכותית

מערכות בינה מלאכותית מסתמכות לעיתים קרובות על מאגרי נתונים גדולים שעשויים להכיל נתונים אישיים. תקנות הגנת המידע, כמו התקנה הכללית להגנת מידע (GDPR) באיחוד האירופי, חלות הן על הפיתוח של מערכות אלה, הן על השימוש בהן. להלן שיקולים רלוונטיים למבקרים בתחום זה:

- מגבלות הנובעות מהמטרה: יש לאסוף נתונים אישיים אך ורק למטרה ברורה ומסוימת. כל עיבוד נוסף חייב להתאים למטרה המקורית, למעט החרגות בהיקף מוגבל. כאשר בינה מלאכותית מאומנת על נתונים היסטוריים (שייתכן כי נאספו לפני תחילת הפרויקט), המטרה המקורית חייבת להצדיק עיבוד נוסף מסוג זה.
- מזעור נתונים: השימוש בנתונים אישיים יוגבל רק לרמה ההכרחית למימוש המטרה שלשמה הם נאספו. האמור חל הן על המודל הסופי, הן על נתונים כלשהם ששימשו במהלך הבדיקה או האימון. על המבקרים לבדוק אילו נתונים היו בשימוש, מהו תפקידם בביצועי המערכת וכיצד הם מטופלים לאורך תהליך הפיתוח.
- מידתיות: חשוב שהכמות והסוג של הנתונים יהיו מידתיים ביחס למטרה ושלא יהיו פולשניים מדי. לדוגמה, זיהוי פנים כדי לעקוב אחר נוכחות תלמידים בבית הספר הוא מן הסתם שימוש מופרז.
- שקיפות: היכולת להסביר תהליכים והחלטות היא דרישה כללית בשירותים לציבור, במיוחד אם נתונים אישיים מעורבים בכך. עיבוד של נתונים אישיים צריך להתבצע בשקיפות.
- הארגון שמפתח את מערכת הבינה המלאכותית ומשתמש בה אחראי לפעול בהתאם לכללי הגנת המידע ולציית ל-GDPR. אם הבינה המלאכותית מעמידה את זכויות הפרט בסיכון גבוה, נדרש לבצע הערכה של ההשפעות על הגנת המידע (DPIA). ה-DPIA צריך להציג ראיות לכך שנשקלו חלופות הכרוכות בסיכונים נמוכים יותר. המבקרים יכולים לבדוק ציות בדרך כלל על ידי עיון בתיעוד.

נספח 3: מדדי שוויון והוגנות במערכות סיווג

הערכת מודלי סיווג מתבצעת בדרך כלל בעזרת שורה של מדדים סטנדרטיים הנגזרים ממטריצת בלבול (confusion matrix) - טבלה המשווה את הניבויים של מודל הסיווג כנגד התוצרים בפועל, כדי למדוד את רמת הדיוק ומדדי ביצוע אחרים הקשורים בכך. חלק ממדדים אלה עשויים לחול על

מודלים של בינה מלאכותית יוצרת, אולם סעיף זה מתייחס ספציפית לבעיות בסיווג.

עם המדדים האלה נמנים:

- **שכיחות:** שיעור המקרים החיוביים בפועל בכל קבוצה.
- **שכיחות מנובאת:** שיעור הניבויים החיוביים בכל קבוצה.
- **שיעור תוצאות חיוביות נכונות (true positive rate - TPR):** שיעור המקרים החיוביים בפועל שזוהו נכונה על ידי המודל.
- **שיעור תוצאות חיוביות שגויות (false positive rate - FPR):** שיעור המקרים השליליים שסווגו באופן שגוי כחיוביים.
- **דיוק (ערך ניבוי חיובי):** שיעור הניבויים החיוביים שזוהו נכונה.
- **ערך ניבוי שלילי (negative predictive value - NPV):** שיעור הניבויים השליליים שזוהו נכונה.

הוגנות ניתנת להערכה בכמה דרכים:

- **יחס מפלה או אי-שוויון הזדמנויות:** הדרישה היא כי שיעורי התוצאות החיוביות הנכונות יהיו דומים עבור קבוצות שונות, בהתמקד על סיכויים שווים לקבל תוצאות חיוביות נכונות.
- **השפעה בלתי שוויונית או צדק חלוקתי:** זו הערכה של מידת הדיוק, והמטרה היא להשיג רמת דיוק דומה של ניבויים חיוביים עבור קבוצות שונות.
- **שוויון סטטיסטי (שוויון דמוגרפי):** הדרישה היא כי שיעור הניבויים החיוביים יהיה זהה עבור כל הקבוצות, ללא קשר להבדלים הקיימים בתוצאות בפועל.

מדדים נוספים של הוגנות כלפי כל הקבוצות:

- **שקילות סיכויים (equalized odds):** הדרישה היא שה-TPR וה-FPR יהיו שווים עבור כל הקבוצות, או שהתראות שווה וזיהויים נכונים יתרחשו בשיעורים דומים עבור כל הקבוצות.
- **דיות (sufficiency):** (שקילות בשיעור הניבוי): הדרישה היא לרמת דיוק שווה וערך ניבוי שלילי שווה עבור כל הקבוצות. לדוגמה, כאשר המודל מנבא שאדם יעבור שוב עבירה, ההסתברות שהמודל צודק זהה עבור כל הקבוצות.

חשוב לציין שבפועל, אף מודל בלתי מושלם לא יכול לעמוד בכל קריטריוני ההוגנות בבת אחת, במיוחד כאשר שיעורי התוצאות החיוביות שונים בבסיסם בין הקבוצות. המבקרים נדרשים להביא בחשבון מגבלות אלה ולהתמקד

בהשפעה הפרקטית של הבדלים כלשהם.

מושגי הוגנות נוספים:

- **הוגנות של תרחישים חלופיים (counterfactual fairness):** הניבוי של המודל לא משתנה אם משנים תכונה רגישה, כאשר כל יתר התנאים שווים.

סקירה טובה של מושגי הוגנות אלה ואחרים מופיעה בספרות⁶⁹. חשוב לציין כי רשימת המדדים הנכללת כאן אינה ממצה, והיא מובאת כנקודת מוצא עבור המבקרים.

- **הוגנות באמצעות חוסר מודעות:** השמטה של תכונות רגישות לא תבטיח הוגנות, מכיוון שמשנתנים אחרים עשויים להימצא במתאם עם תכונות אלה.
- **הוגנות אישית:** אנשים דומים אמורים לקבל ניבויים דומים, אם כי פרמטר זה קשה בדרך כלל למדידה.

נספח 4: מילון מונחים

הדירות (Reproducibility)

הדירות היא היכולת לשחזר תוצאות כשמשתמשים באותם נתונים, אותם פרמטרים ואותה סביבה.

הוגנות (מובנית בתכנון) (Fairness (by design))

הוגנות פירושה יחס שוויוני לפרטים ולקבוצות והימנעות מהשפעות שליליות בשל מאפיינים כמו מגדר, מוצא או מיקום. פלט לא הוגן של מערכת עלול לגרום לאפליה, להפסד כספי או לנחיתות חברתית בדרך אחרת. הוגנות מובנית בתכנון היא תפיסה המטמיעה הוגנות מהרגע הראשון, במהלך שלב התכנון והפיתוח של מערכת בינה מלאכותית, במקום תיקון בדיעבד.

הזיה (Hallucination)

הזיה של בינה מלאכותית היא מצב שבו מערכת מפיקה פלט שמכיל מידע שגוי או שאינו מבוסס על עובדות.

היפר-פרמטר (Hyperparameter)

היפר-פרמטרים משמשים לשליטה בארכיטקטורת המודל ובתהליך הלמידה בלמידת מכונה. פרמטרים אלה נשלטים על ידי כותבי הקוד, בשונה מערכי הפרמטרים של המודל, הנוצרים באמצעות אימון.

הנדסת מאפיינים (Feature Engineering)

הנדסת מאפיינים היא התהליך של הפיכת נתונים גולמיים למשתנים בעלי משמעות עבור המודל.

הסבריות (מובנית בתכנון) (Explainability (by design))

הסבריות היא היכולת לתאר כיצד המערכת מגיעה לתוצרים או להחלטות. הסבריות מובנית בתכנון היא מתודולוגיה הדורשת להבטיח הסבריות מהרגע הראשון, במהלך התכנון והפיתוח של מערכת בינה מלאכותית.

מילון המונחים מוסיף פרטים על המינוח המשמש לאורך מסמך זה. הודגשו במיוחד המונחים הטכניים הנוגעים לאלגוריתמים ובינה מלאכותית.

בינה מלאכותית (Artificial Intelligence - AI)

מערכות בינה מלאכותית הן טכנולוגיות מעשה ידי אדם שייעודן לבצע מטלות אשר באופן טיפוסי מצריכות אינטליגנציה (בינה). מערכות אלה לומדות מתוך נתונים ומתאימות את עצמן להשגת מטרות ספציפיות, כך שעל פי רוב הן משפרות את ביצועיהן בעצמן, ללא תכנות נוסף במפורש. בינה מלאכותית כוללת תחומים כמו למידת מכונה, עיבוד שפה טבעית, ראייה ממוחשבת, זיהוי קולי, תכנון ורובוטיקה. היא מאפשרת יכולות כמו תפיסה, הנמקה, קבלת החלטות ופתרון בעיות. מרבית מערכות הבינה המלאכותית מסתמכות על למידת מכונה, שבמסגרתה מודלים לומדים דפוסים מתוך נתונים כדי להשיג מטרות ספציפיות.

בינה מלאכותית יוצרת (Generative AI)

מערכות בינה מלאכותית יוצרת יודעות ליצור תוכן חדש (כמו טקסט, תמונות, וידאו או אודיו) על ידי למידה מתוך קובצי נתונים גדולים. הן מסוגלות לייצר תוצרים ריאליסטיים, על פי רוב עם הכוונה אנושית מעטה או ללא הכוונה כזו כלל.

בינה מלאכותית לחיזוי (Predictive AI)

מערכות בינה מלאכותית לחיזוי משתמשות בנתונים היסטוריים כדי לזהות דפוסים ולספק ניבויים לגבי מקרים חדשים. הן מסוגלות להעריך את הסיכון להתפתחות מחלה על פי רשומות רפואיות או לסמן טעויות פוטנציאליות בדוחות מס. מערכות אלה משמשות באופן שכיח לצורכי סיווג (כמו מיון מקרים לקטגוריות) או גרסיה (כמו ניבוי ערך מספרי).

69 ראו לדוגמה: S. Verma & J. Rubin (2018): [Fairness Definitions Explained](#)

הרעלת נתונים (Data Poisoning)

הרעלת נתונים היא הוספת דוגמאות שנעשו בהן מניפולציות לנתוני האימון, כדי לפגוע בביצועי מערכת הבינה המלאכותית או לשנות את התנהגותה בדרך ספציפית.

התאמת יתר (Overfitting)

התאמת יתר היא מצב שבו מערכת בינה מלאכותית מתאימה את עצמה מדי לנתוני האימון, ובעקבות זאת היא לא מספקת ביצועים טובים על נתונים חדשים.

יצירה מבוססת אחזור (Retrieval Augmented Generation - RAG)

יצירה מבוססת אחזור היא גישה המאפשרת למודלי שפה גדולים לשאוב ממקורות מידע מהימנים כשהם יוצרים תשובה. הזנת מסמכים או נתונים רלוונטיים לתוך ההקשר של המודל מאפשרת ל-RAG להתמודד עם אתגרים ידועים בבינה מלאכותית יוצרת, כמו הזיות, מידע מיושן והיעדר ציטוטי מקורות.

כוונון עדין (Fine-tuning)

כוונון עדין הוא תהליך שבו מכניסים התאמות למודל שפה גדול (LLM) באמצעות דוגמאות ספציפיות, כדי שתשובותיו יהיו מדויקות יותר עבור נושאים מסוימים.

למידה עמוקה (Deep Learning)

בלמידה עמוקה, מערכות בינה מלאכותית משלבות כמה שכבות עיבוד כדי לפתור משימות מורכבות.

מהימנות (Reliability)

מהימנות מתייחסת ליכולת של המודל לספק ביצועים עקביים ונכונים בתנאים הצפויים, בפרט בעת טיפול בנתונים הדומים לאלה ששימשו לאימון המודל.

מודלי יסוד (Foundation Models) או מודלי בינה מלאכותית למטרות כלליות (General-purpose AI Models)

מודלי יסוד או מודלי בינה מלאכותית למטרות כלליות מאומנים על מאגרי נתונים גדולים ומגוונים מאוד. הם משתמשים בטכניקות למידה עמוקה, וניתן להתאים אותם למגוון רחב של יישומים, מזיהוי תמונות ועד מחקר מדעי.

מודלי שפה גדולים (Large Language Models - LLMs)

מודלי שפה גדולים הם סוג עיקרי של מודלי יסוד, המאומנים בעיקר על טקסטים. הם מסוגלים לבצע מטלות רבות, כמו תרגום, סיכום ושיחה, על ידי הפקת שפה הדומה לשפה אנושית.

נתוני אימון (training), אימות (validation) ובדיקה (test)

בתחילה משתמשים בנתוני אימון כדי לאמן מודל למידת מכונה לביצועים מיטביים. לאחר מכן משתמשים בנתוני אימות כדי להעריך את התאמת המודל תוך כדי כונון ההיפר-פרמטרים. נתוני הבדיקה הם האחרונים, ומטרתם לתת הערכה סופית של ביצועי המודל.

עמידות (Robustness)

עמידות היא היכולת של המודל לשמור על רמת ביצועיו כאשר הוא עומד בפני קלט בלתי צפוי - כמו נתונים בהתפלגות שונה מהנתונים שעליהם הוא אומן; נתונים "מחוץ להתפלגות" (out of distribution - OOD); רעש; או שינויים בסביבה.

פרומפט (Prompt)

פרומפט (או הנחיה) הוא קלט משתמש שניתן למערכת בינה מלאכותית יוצרת. מונח זה משמש בעיקר בהקשר של LLM.

פרטיות (מובנית בתכנון) (Privacy (by design))

פרטיות כוללת הגנה על כל הנתונים האישיים שמערכת הבינה המלאכותית אוספת, משתמשת בהם, משתפת או מאחסנת. פרטיות מובנית בתכנון היא תפיסה המטמיעה פרטיות מהרגע הראשון, במהלך שלב התכנון והפיתוח של מערכת בינה מלאכותית, במקום תיקון בדיעבד.

שקיפות (מובנית בתכנון) (Transparency (by design))

שקיפות משמעה מתן מידע ברור על השימוש במערכת - איך, מתי ומדוע משתמשים בה. שקיפות מובנית בתכנון היא מתודולוגיה הדורשת להבטיח שקיפות במהלך כל שלב בתהליך התכנון והפיתוח של מערכת בינה מלאכותית.

תפעול למידת מכונה (Machine Learning Operations - MLOps)

MLOps הוא מכלול של פרקטיקות המסייעות לנהל, לבדוק ולעדכן מודלים של למידת מכונה.

Cross-Industry Standard Process for Data Mining ((CRISP-DM

CRISP-DM הוא תהליך תקני לתכנון ופיתוח של מערכות למידת מכונה ובינה מלאכותית. הוא מורכב משישה שלבים:

- הבנה עסקית
- הבנת הנתונים
- הכנת הנתונים
- בניית מודל
- הערכה
- פריסה

שתפקיד היועץ הסתיים, צריך להיות כוח אדם פנימי שרכש את הידע לגבי התפקיד הרלוונטי. לפיכך, באחריות המבוקר להבטיח תיעוד מספק של כל עבודה המתבצעת על ידי היועצים החיצוניים.

ארכיטקט/מהנדס בינה מלאכותית

אדם זה מגדיר את הארכיטקטורה הכוללת של מערכת בינה מלאכותית ואת אסטרטגיית השילוב שלה. הוא מסייע להגדיר כיצד מודלים, צנרות נתונים, APIs ויישומים ישתלבו יחד. הוא בוחר מסגרות, פלטפורמות ושירותי ענן לפריסת המערכת, ודואג להיבטים של סקלריות ואבטחה.

בעלי הפרויקט/בעלי המוצר

הצוות או היחידה בארגון המבוקר האחראים למשימה שמעתה תיתמך במערכת בינה מלאכותית או תבוצע אוטומטית בעזרתה. הם מחליטים אילו מדדי ביצוע נדרשים ממערכת הבינה המלאכותית ומה הם תנאי הקבלה של התוצרים בסיום פרויקט הפיתוח של המערכת.

חשב

האדם שעורך ביקורת של פרויקטים ובודק עמידה בעקרונות ממשל.

אנליסט/מדען נתונים

האדם שעובד עם הנתונים שיוזנו לתוך מערכת הבינה המלאכותית ומנתח אותם. הוא אחראי להבין את הנתונים, ועליו להיות מעורב באופן מעמיק בתהליך הפיתוח. הוא מסייע לבעלים של המוצר, על ידי תרגום דרישותיו למפרטים ודרישות שיועברו למפתחים.

מהנדס נתונים

האדם האחראי על ההיבטים הטכניים של הנתונים הגולמיים (מחסי נתונים, איכות נתונים, בקרת גישה) וכן על הבנת הנתונים הגולמיים ומקורותיהם. הוא גם אחראי על אספקת הנתונים וניהול הנתונים.

מומחה לתחום העיסוק (Subject matter expert)

מונח גנרי המתייחס לאדם בעל ידע מומחה בתחום (domain) ספציפי.

AI Artificial Intelligence

API Application Programming Interface

BSI German Federal Office for Information Security

CRISP-DM Cross-Industry Standard Process for Data Mining⁷⁰

DPIA Data Protection Impact Assessment

FLOP Floating Point Operation

GDPR General Data Protection Regulation⁷¹

GPU Graphics Processing Unit

ICO Information Commissioner's Office

LGPD Brazil's General Data Protection Law

LLM Large Language Model

ML Machine Learning

NIST U.S. National Institute of Standards and Technology

OECD Organisation for Economic Co-operation and Development

RAG Retrieval Augmented Generation

SAI Supreme Audit Institution

TCU Brazil's Federal Court of Accounts

נספח 5: תפקידים

ברשימה הבאה ננסה לסכם את אחריותם של בעלי תפקידים בפרויקט בינה מלאכותית שעשויים להיות רלוונטיים לביקורת. המטרה היא לסייע למבקרים לזהות אנשי קשר בארגון המבוקר. תפקידים בתחום הבינה המלאכותית משתנים ככל שהטכנולוגיות וגישות העבודה מבשילות, כך שאין זו רשימה ממצה, וקרוב לוודאי שתשתנה עם הזמן כדי לשקף כלים, מתודולוגיות ושיקולי אתיקה חדשים.

ארגונים מבוקרים קטנים יחסית עשויים לשלב כמה מהתפקידים הנמנים למטה באדם יחיד או בצוות יחיד.

חלק מהתפקידים יכולים גם להתבצע על ידי יועצים חיצוניים; אולם במקרה שהביקורת מתקיימת לאחר

A. Clark (2018): The Machine Learning Audit-CRISP-DM Framework 70

71 הפרלמנט האירופי ומועצת האיחוד האירופי (2016): תקנה (האיחוד האירופי) 2016/679 של הפרלמנט האירופי של המועצה מיום 27 באפריל 2016 בעניין הגנה על אישיות טבעית ביחס לעיבוד נתונים אישיים ובעניין תנועה חופשית של נתונים כאמור, וביטול דירקטיבה EC/95/46 (התקנה הכללית להגנת מידע).

ממונה על אבטחת ID

זהו האדם הבכיר ביותר האחראי על אבטחת טכנולוגיית המידע בארגון המבוקר. עליו להכיר את כל היבטי האבטחה של הפיתוח והתפעול של מערכת הבינה המלאכותית.

ממונה על הגנת מידע ופרטיות

זהו האדם הבכיר ביותר האחראי על הגנת מידע בארגון המבוקר. עליו להיות מודע לכל חשש שעולה לגבי שימוש בנתונים אישיים על ידי מערכת הבינה המלאכותית. תפקידו לוודא שהתוכנה מצייתת להוראות דיני פרטיות המידע והתקנות בתחום זה, כדוגמת ה-GDPR באיחוד האירופי.

ממונה על התקציב

אדם זה אחראי לתקציב של הארגון המבוקר, ועל כן אחראי על כל הוצאה הקשורה לפרויקטים של פיתוח מערכות בינה מלאכותית, רכש שלהן או ייעוץ לגביהן. בידיו הסמכות לקבוע אם הפיתוח והתפעול של תוכנה כזו מייצגים ניצול משתלם של תקציב הארגון המבוקר, ועליו להחזיק בכל המידע התקציבי על הפיתוח והתפעול של המערכת.

מנהל הפרויקט

האדם האחראי לכל נושאי הניהול/ממשל של הפרויקט. עליו להיות מסוגל לספק את כל המסמכים הנדרשים הקשורים לניהול הפרויקט.

מפתח

האדם או האנשים שמייצרים את מערכת הבינה המלאכותית בהתאם למפרטים ולדרישות שהוסכמו עם בעלי המוצר (ומאמנים את המודל, עבור מודלים שנדרש בהם שלב אימון). המפתחים אחראים להפיכת הנתונים הגולמיים למשתנים הסופיים שבהם ישתמש המודל ("הנדסת מאפיינים"), והם עובדים בשיתוף עם מדעני הנתונים ומהנדסי הנתונים, מנהל הפרויקט ובעלי המוצר.

משתמש/אחראי עיבוד

האדם או היחידה שאמורים להשתמש בתוצאות מערכת הבינה המלאכותית לצורך מילוי תפקידם בארגון. הם גורם מכריע בקביעת הצלחתם של פרויקטים, מכיוון שעליהם להבין את ההצעות או התוצאות ממערכת הבינה המלאכותית וליישם אותן במשימות (שגרתיות).

סמנכ"ל מערכות מידע (CIO)

ה-CIO של הארגון המבוקר אחראי לכל מערכות טכנולוגיית המידע של הארגון, ועל כן עליו להכיר את כל מערכות הבינה המלאכותית שכבר פועלות ואת כל הפרויקטים העוסקים בפיתוח תוכנה מסוג זה.

קו חם לתהליכי עיבוד/מוקד תמיכה למשתמשים

הצוות המופקד על מתן תמיכה למשתמשים או לאנשים העוסקים בעיבוד. עליו לדעת להשיב על כל השאלות המתעוררות במהלך התפעול השגנתי של התוכנה.

כלי עזר לביקורת

כלי העזר לביקורת הוא כלי אקסל המכיל שאלות ביקורת מפורטות וראיות ביקורת שעל המבקרים לחפש בעת ביצוע ביקורת בינה מלאכותית. השאלות מסודרות בהתאם לשלבים השונים במחזור החיים של המערכת.

המבקרים יכולים לאסוף שאלות רלוונטיות לביקורת הספציפית וליצור שאלון מותאם לצורכיהם. בדרך זו, הכלי יכול לסייע להכין ולהנחות את המבקרים לאורך שלבי הביקורת של מערכות בינה מלאכותית. הכלי נועד לסייע כעמדת מוצא, ונועד להקיף את ההיבטים החשובים ביותר המתוארים במדריך זה.

[הורידו כאן את כלי העזר לביקורת](#)



בינה מלאכותית וביקורת ציבורית: אתגרים והזדמנויות עבור מוסדות ביקורת עליונים

דולורס גנארו-מויה, אנטוניו מנאל לופז-הרננדז ומריה גודז^{1,2}

זהו תרגום של המאמר Artificial Intelligence and Public Sector Auditing: Challenges and Opportunities for Supreme Audit Institutions. המאמר פורסם במקור בכתב העת <https://www.mdpi.com/2673-4060/6/2/78>, 78, 6(2), 2025, World.

1 ד"ר דולורס גנארו-מויה - המחלקה לכלכלה בין-לאומית וספרדית, הפקולטה ללימודי כלכלה ועסקים, אוניברסיטת גרנדה, 18071 גרנדה, ספרד, lolagenaro@gmail.com; ד"ר אנטוניו מנאל לופז-הרננדז - המחלקה לחשבונאות ומימון, הפקולטה ללימודי כלכלה ועסקים, אוניברסיטת גרנדה, 18071 גרנדה, ספרד; ד"ר מריה גודז - המחלקה למדעי המחשב ובינה מלאכותית, בית הספר להנדסת מחשבים ותקשורת, אוניברסיטת גרנדה, 18014 גרנדה, ספרד, mariagodz@correo.ugr.es. עורכים אקדמיים: סוטיריוס ברסימיס, אנדראס פאוסטרס, דימיטריוס א' גאורגקלוס ואורטווין רן. זכויות יוצרים: © 2025 - המחברים. בעל הרישיון: MDPI, באזל, שווייץ. מאמר זה הוא מאמר בגישה פתוחה המופץ על פי תנאי רישיון Creative Commons Attribution (CC BY) (<https://creativecommons.org/licenses/by/4.0/>).

2 עורכים אקדמיים: סוטיריוס ברסימיס, אנדראס פאוסטרס, דימיטריוס א' גאורגקלוס ואורטווין רן. זכויות יוצרים: © 2025 - המחברים. בעל הרישיון: MDPI, באזל, שווייץ. מאמר זה הוא מאמר בגישה פתוחה המופץ על פי תנאי רישיון Creative Commons Attribution (CC BY) (<https://creativecommons.org/licenses/by/4.0/>).

המאמר בקצרה

- השימוש בבינה מלאכותית (artificial intelligence - AI) בגופים ציבוריים גדל בצעדי ענק ותורם לשיפור התכנון ומתן השירותים וכן לניהולם הפנימי וליעילותם של מוסדות ציבור, אולם הפוטנציאל הטמון בה כרוך גם בשורה של סיכונים.
- כדי לבצע ביקורת בשירותים ציבוריים ובמדיניות ציבורית, מוסדות ביקורת עליונים (Supreme Audit Institutions - SAI) נדרשים להתאים את משאביהם האנושיים והטכנולוגיים לתרחיש חדש זה.
- מאמר זה מנתח את השלכות חדירתה של הבינה המלאכותית למגזר הציבורי וכן את האתגרים שהתפתחויות טכנולוגיות אלה מציבות בפני מוסדות ביקורת עליונים הפועלים לשיפור האפקטיביות והיעילות של משימות הביקורת שלהם.
- המאמר מציג ניתוח תפיסתי ומחקרי המבוסס על ראיות תיעודיות וחקר מקרים.
- עבודה זו גם עוסקת בביקורת המתמקדת במערכות מבוססות אלגוריתמים שמבצעים חלק ממוסדות הביקורת העליונים.

הקדמה | בינה מלאכותית במינהל הציבורי | בקרה חיצונית על ביצועי המגזר הציבורי לאור השימוש בבינה מלאכותית | מסקנות

1. הקדמה

הזדמנויות לשיפור האפקטיביות של הביקורת ויעילותה.

בהקשר זה, מטרת המאמר היא כפולה. לאור התפתחותה המהירה והמתמשכת של טכנולוגיית הבינה המלאכותית, מטרת המאמר היא לאמץ גישה מושגית וחוקרת. אף שהניתוח כאן מתבסס על תיעוד אמפירי קיים וחקר מקרים, יש לראות בו תוצר שנועד לפרשנות והדגמה ולא מסקנה סופית. המטרה היא לעודד חשיבה מוסדית ולספק בסיס לנושאי מחקר עתידיים. ראשית, העבודה מבקשת למלא את הפער הקיים בספרות האקדמית והמוסדית בנוגע להשפעת השימוש הגובר בבינה מלאכותית על בקרה חיצונית של המינהל הציבורי, בפרט מנקודת המבט של מוסדות הביקורת העליונים. אף שמספר הולך וגדל של מחקרים בוחן את היישום של בינה מלאכותית בשירות הציבורי, חסר ניתוח המראה כיצד גופי פיקוח חיצוני מסתגלים לשינוי הפרדיגמה הזה - במיוחד בספרד ובתחומי שיפוט אחרים בעלי מבנה ביקורת דומה. שנית, מאמר זה שואף לנתח את ההזדמנויות והאתגרים אשר הפיתוח והיישום של בינה מלאכותית מניחים לפתחם של מוסדות ביקורת עליונים. כדי לעשות זאת, אנו מבצעים ניתוח איכותי ותיעודי המבוסס על דוחות עדכניים, יוזמות ביקורת ומסגרות אסטרטגיות ממוסדות ביקורת עליונים של מדינות שונות. פרק ספציפי מוקדש להצגת ניתוח מצב הנוכחי של פרקטיקות ביקורת הקשורות לבינה מלאכותית, והוא כולל זיהוי גישות מתפתחות, התאמות טכנולוגיות וחסמים מוסדיים. אבחון זה יכול לשמש כתמונת מצב התחלתית (baseline) להערכת ההתפתחויות העתידיות בתחום.

עבודה זו מיישמת גישה איכותנית-תיעודית, הבוחנת דוחות רשמיים של מוסדות ביקורת עליונים ומאמרים בגישה פתוחה. המסגרת האנליטית שלנו מתמקדת במשימות הבאות: 1. זיהוי פרקטיקות מתפתחות וגישות מומלצות ביישום טכנולוגיות בינה מלאכותית בביקורת ציבורית; 2. מגבלות טכניות; 3. כשירות מוסדית. גישה זו נבחרה על רקע המחקר האקדמי המצומצם על אודות שילוב בינה מלאכותית בביקורת ציבורית של מוסדות ביקורת עליונים, כדי לזהות נושאים קריטיים ולהנחות מחקר תיאורטי עתידי.

2. בינה מלאכותית במינהל הציבורי

כניסתה של הבינה המלאכותית לשימוש בכל תחומי החיים מתבררת, בנקודה זו במאה ה-21, כמהפכה גדולה הדומה בהיקפה למהפכה התעשייתית הראשונה במאה ה-18, למהפכה השנייה שהתרחשה עם כניסת הייצור ההמוני או הסדרתי במאה ה-19 או למהפכה הטכנולוגית השלישית של המאה ה-20², ומכאן קביעת שמה למהפכה התעשייתית הרביעית (Industry 4.0). הבינה המלאכותית צמחה אומנם באמצע המאה הקודמת, אך כוח העיבוד הגדול של המחשבים לצד התרחבות השימוש היום-יומי

אופיו הדינמי והמהיר של המחקר בתחום הבינה המלאכותית מבליט עד כמה חשוב לעדכן מסגרות והנחות כאשר מחקרים אמפיריים חדשים וכיתוחים טכנולוגיים באים לעולם. בהקשר זה, השימוש בבינה מלאכותית משפיע ישירות על המגזר הציבורי, בפרט בכל הקשור לתפקידו של המגזר הציבורי בפיקוח ואסדרה (רגולציה). מו-אהל¹ מציין כי יש לצפות שהבינה המלאכותית תהיה נוכחת בכל ארגוני המגזר הציבורי חרף חסמי הכניסה הגבוהים הנובעים מעלותה הגבוהה, וזאת בשל השפעתה על היעילות בטווח ארוך, שתעלה ככל שהאלגוריתמים ישתפרו באמצעות עיבוד נתונים.

התפתחותה של הבינה המלאכותית עומדת להציב בשנים הקרובות אתגר בפני מוסדות ביקורת עליונים - המוסדות האמונים על בקרה עצמאית וחיצונית על הפעילות הכלכלית והפיננסית של המגזר הציבורי.

מוסדות ביקורת עליונים ממלאים תפקיד מכריע בכל מדינה דמוקרטית, שכן הם העוקבים אחר השימוש ההולך בכספי ציבור ומפקחים על החוקיות, הדיוק, היעילות והאפקטיביות של פעילות המגזר הציבורי. סוג הבקרה שהם מפעילים הוא חיצוני, בניגוד לבקרה פנימית או מקדימה בתוככי הגופים הציבוריים, ועל כן זוהי בקרה שאינה תלויה בגוף המבוקר. תוצאות עבודתם של מוסדות הביקורת נשלחות אל המחקק כדי שיוכל להשתמש בהן לצורך בקרה על הממשלה. כדי למלא את תפקידם לבקר, מוסדות הביקורת מיישמים טכניקות, שיטות ותקנים של ביקורת שיושפעו מכניסתן של טכנולוגיות חדשות.

במובן זה, כניסתה של הבינה המלאכותית היא כבר עתה גורם חשוב אשר מניע שינוי בעבודת הביקורת, הן במגזר הציבורי והן בפרטי. משום כך יש לצפות לשינוי מהותי באופן שבו מוסדות ביקורת עליונים יבצעו את תפקידי הפיקוח והבקרה המוטלים עליהם. סביר לשער כי יהיה עליהם להתמודד עם הצורך להתאים את משאביהם האנושיים והטכנולוגיים למציאות שהולכת ונעשית מורכבת יותר. כדי לעשות זאת, יהיה עליהם להכין אסטרטגיות ולקבל החלטות שיאפשרו להיערך לשינוי בטווח הקצר והארוך.

במספר מדינות, מוסד הביקורת הלאומי מוסמך גם לפעול כערכאה שיפוטית שמטרתה העיקרית היא להשיב לקופת המדינה כספי ציבור שנוצלו שלא כדין. כך נהוג לדוגמה בספרד, צרפת ומרוקו. פונקציה נוספת זו של מוסד הביקורת תושפע קרוב לוודאי מכניסת הבינה המלאכותית, אם כי המאמר הנוכחי אינו מתמקד בתחום זה.

מכל מקום, השימוש בבינה מלאכותית להפעלת בקרה חיצונית על הניהול הכלכלי והפיננסי של המגזר הציבורי, כמו בשלל תחומים אחרים, עתיד לפתוח מגוון של

מעבר להגדרות שצוינו, קיימות דרכים מגוונות לסווג בינה מלאכותית. אחת הגישות המקובלות ביותר מבוססת על היכולת התפקודית ומבחינה בין בינה מלאכותית יישומית - הידועה גם כבינה מלאכותית צרה או חלשה, לבין בינה מלאכותית כללית ובינה מלאכותית חזקה¹⁰. הסוג הראשון של בינה מלאכותית כולל מערכות שמבצעות מטלות ספציפיות, כמו זיהוי קולי, אבחונים רפואיים או בקרת רובוטים, אך אין בכוחן להבין את העיקרון הכללי וליישם אותו על מטלות אחרות. מחשבים מבצעים מטלות באופן מושכל בתחומים ספציפיים, באופן שתואם יכולות אנושיות שונות¹². אומנם למערכות בינה מלאכותית חלשות יש יישומים מוגבלים, אך בזכות פשטותן הן מועילות במיוחד למטלות הכוללות מספר תנועות חזרתיות ומספר משתנים. בהתייחס למצב פעולה זה, ניתן לזהות סוגים שונים של בינה מלאכותית - כמו עיבוד שפה טבעית, למידת מכונה וראייה מלאכותית¹³. בינה מלאכותית כללית מתייחסת לרעיון שבינה מלאכותית יכולה להפגין באמצעות מכונות יכולות אנושיות¹⁴, ואף שהיא עדיין אינה קיימת - היא יעד מחקר בתחום הבינה המלאכותית. מאידך גיסא, בינה מלאכותית חזקה תציג יכולות גבוהות יותר מאלה של בני אדם, ומכונות יספקו ביצועים טובים יותר מביצועים אנושיים ויעקפו אותם. משום כך, היא נחשבת בגדר איום פוטנציאלי על החברה.

כלי בינה מלאכותית מסווגים על פי יכולותיהם, כדלהלן¹⁵:

1. מכונות תגובתיות (Reactive Machines): מערכות בינה מלאכותית שעובדות על נתונים זמניים ומגיבות לגירויים חיצוניים בזמן אמת, ללא יכולת לשמור נתונים או ללמוד מחוויות קודמות. "כחול עמוק" (DeepBlue - מחשב-על שהביס את אלוף העולם בשחמט קספרוב בשנות ה-90) או מנוע ההמלצות של נטפליקס הם שתי דוגמאות לבינה מלאכותית מסוג זה.
 2. זיכרון מוגבל (Limited Memory): מערכות בינה מלאכותית שמסוגלות לשמור חוויות קודמות ולהשתמש בהן כדי לתת תחזיות או החלטות, אך קיבולת הזיכרון שלהן מוגבלת. דוגמאות לסוג זה הן ChatGPT 4, סיינעים וירטואליים, צ'אטבוטים כמו סירי ואלקסה ומכונות אוטונומיות.
 3. תיאוריית התודעה (Theory of Mind): מערכות בינה מלאכותית שמסוגלות להבין היבטים פסיכולוגיים ורגשיים של בני אדם ולהגיב אליהם - יכולת שעשויה להוביל לאינטראקציות טבעיות ואינטואיטיביות יותר.
 4. מודעות עצמית (Self-Awareness): מערכות בינה מלאכותית שמסוגלות לחשוב ולפעול. הן בעלות מודעות עצמית והבנה עצמית ויכולו לחשוב על הקיום ועל מערכות הערכים של עצמן.
- שני הסוגים האחרונים אומנם אינם קיימים עדיין, אך מדענים עובדים על פיתוחם.

בטלפונים ניידים ובמכשירים טכנולוגיים המבוססים על מרשתת הדברים (IoT) מספקים כמות גדולה של מידע מובנה ובלתי מובנה. מידע זה מעובד ומשמש ל"הזנה" של האלגוריתמים שהם הבסיס לבינה מלאכותית. כצפוי, השימוש בכלים מבוססי בינה מלאכותית התפשט בזירה הפרטית והציבורית, אם כי במאמר זה נתמקד רק בזו האחרונה.

כצעד מקדים לניתוח השלכות הבינה המלאכותית במגזר הציבורי, יש מקום לסקירה כללית של טכנולוגיה זו. להלן נסקור בקצרה את המבנה התפיסתי והטיפולוגי של הטכנולוגיה, כדי לבסס מסגרת שיטתית שתאפשר הערכה טובה יותר של השפעותיה על גופים ציבוריים.

2.1 תפיסת היסוד של הבינה המלאכותית וסוגי בינה מלאכותית

בספרות האקדמית אין הגדרה מוסכמת של בינה מלאכותית³. את המונח טבע בשנות ה-50 של המאה העשרים מקארתי⁴ שהגדיר אותו כ"מדע של ייצור מכונות נבונות". בזמנו היה מדובר רק בשימוש באותן מכונות לצורך הבנת התבונה האנושית, בלי הגבלה רק לשיטות המחקות תהליכים ביולוגיים כפי שהם נצפים בטבע. השימוש הראשון במונח זה בהקשר האקדמי ציין תחום מתפתח במחקר, שחקר מחד גיסא את יכולתן של מכונות לבצע משימות המשקפות התנהגות מושכלת הדומה להתנהגות אנושית, ומאידך גיסא את יכולתן של מכונות להתנהג כסוכנים נבונים בעלי תפיסה סביבתית שמבצעים פעולות להשגת מטרות מסוימות⁵. בהמשך, ההגדרה שונתה כדי להתאימה לאופייה המשתנה של הבינה המלאכותית. לדוגמה, לדידו של רייץ⁶, בינה מלאכותית היא מחקר הבודק כיצד לגרום למחשבים לבצע דברים שבעת הזו אנשים עושים טוב מהם, והוא מתמקד במחשבים כמכשירי בינה מלאכותית. הבינה המלאכותית הוגדרה גם בידי פטרסון⁷ - כענף במדעי המחשב הנוגע לחקר ויצירה של מערכות מחשב המפגינות צורה כלשהי של בינה, מערכות הלומדות מושגים וטעמים חדשים, מערכות שמסוגלות לחשיבה הגיונית ולהסקת מסקנות מועילות על העולם שסביבן, מערכות שמסוגלות להבין שפה טבעית או לתפוס ולהבין סצנה ויזואלית ומערכות שמבצעות סוגים אחרים של משימות הדורשות סוגים אנושיים של בינה.

בשנים האחרונות הגדיר ה-OECD⁸ מערכות בינה מלאכותית כמערכות מבוססות מכונה אשר מסוגלות לספק עבור שורה נתונה של מטרות המוגדרות בידי אדם תחזיות, המלצות או החלטות המשפיעות על סביבות אמיתיות או וירטואליות. מערכות בינה מלאכותית נועדו לפעול ברמות משתנות של אוטונומיה. נוסף על כך, הבינה המלאכותית כוללת מכונות שמבצעות משימות קוגניטיביות כמו שמבצעים בני אדם.

שכיח גם לסווג בינה מלאכותית כסימבולית או כמבוססת-כללים ולא-סימבולית. בינה מלאכותית סימבולית, הידועה גם כ"מערכת מומחית" ("expert system") משתמשת בסמלים ובכללים לוגיים כדי לייצג מידע ולבצע בו פעולות (מניפולציות), לתאר זרימות עבודה ולהפיק תוצאות. בזכות פשטותה היחסית, בינה מלאכותית מסוג זה מתאימה יותר לתהליכים או בעיות המתאפיינים במורכבות נמוכה ובהשתתפות מספר מועט של שחקנים, שבהם הפעולות המבוצעות הן מעטות והשינויים אינם תדירים¹¹. בינה מלאכותית לא-סימבולית, המתייחסת למידת מכונה (machine learning - ML), מורכבת מסדרה של טכניקות המאפשרות למכונות ללמוד ולספק ניבויים מתוך נתונים היסטוריים על בסיס זיהוי דפוסים, ללא צורך בהוראות של אדם.

בשנים האחרונות, למידת מכונה היא המוקד הבולט בפיתוח בינה מלאכותית, והיא מאפשרת לה ללמוד ולשפר את פעולותיה בהתבסס על נתונים, ללא תכנות ספציפי. מערכות למידת מכונה כוללות רשתות עצביות מלאכותיות (artificial neural networks - ANN), ובתוך אלה, מערכות למידה עמוקה (deep learning). כולן מייצגות ביטויים ספציפיים או תתי-קבוצות של מערכות בינה מלאכותית צרות ובעלות זיכרון מוגבל^{16,17} כמתואר לעיל.

2.2 חדירת הבינה המלאכותית למגזר הציבורי: הזדמנויות וסיכונים

אף שהדיונים על הפוטנציאל של הבינה המלאכותית מתמקדים בעיקר במגזר הפרטי, גופים ציבוריים תופסים בהדרגה מקום בולט בתחום הבינה המלאכותית וממלאים מגוון רחב של תפקידים, לרוב במקביל. לדוגמה הם עשויים לשמש גורמי מימון ומשקיעים ישירים, רוכשי פתרונות קיימים, רגולטורים, מגשרים, מנהלי נתונים או משתמשי נתונים ונותני שירותים¹⁸.

ככל שתפקידיו של המגזר הציבורי ימשיכו להתפתח, כוחה החדשני ומשנה המציאות של הבינה המלאכותית ייצור קרוב לוודאי עוד ועוד הזדמנויות, שיתמקדו בעיקר בתחומים הבאים¹⁹:

1. שיפור קבלת ההחלטות במינהל הציבורי, ובמיוחד גיבוש מדיניות ציבורית, יישומה והערכתה: מערכות בינה מלאכותית יכולות לשקף את האינטרסים והנושאים המעסיקים את האזרחים, לזהות מגמות ולצפות מראש מצבים הדורשים את תשומת ליבם של גופים ציבוריים. המטרה היא לחזות מראש תוצאות או השפעות אפשריות ובדרך זו להגדיל את הסיכויים להתערבויות מוצלחות²⁰;

2. שיפור בתכנון והענקת שירותים כוללניים יותר לאזרחים ולעסקים: איסוף ועיבוד של נתונים דיגיטליים המקדמים שיפור בשירותים לציבור, מאפשרים מעורבות אינטראקטיבית מול הציבור, מציעים הכוונה או מעבירים מידע חיוני. כל אלה תורמים להשתתפות האזרח בפעילות המגזר הציבורי. בתחום התשתיות - יישומי בינה מלאכותית מאפשרים תחזוקה מונעת, תיקון כשלים או תכנון זמני שימוש בתשתיות בהתאם לביקוש, דבר התורם לניצול משאבים יעיל יותר²¹;

3. שיפור הניהול והיעילות הפנימית של מוסדות ציבור: מערכות בינה מלאכותית יכולות לסייע בביצוע מטלות שבאחריות העובדים. כך הן משחררות עובדי ציבור ממטלות שגרתיות ומפנות עבורם זמן לעסוק בפעילויות שיש בהן יותר ערך ומורכבות. הן יכולות גם לסייע בהקצאת משאבים כספיים וניהולם באמצעות זיהוי ומניעה של מרמה וניתוב כספים שלא כדון, זיהוי חוסר יעילות בהקצאת כספי ציבור וניצולם ואיתור בעיות נוספות. בהקשר זה, סוילו ואחרים²³ מציעים לראות בממשלת סלובניה מקרה מבחן. שם התגלו הודות לזמינותם של נתונים פתוחים על רכש ציבורי חריגות שהתבטאו בזינוק ברמת ההוצאות בתקופות משבר, בתקופות בחירות או בעת מגפת הקורונה העולמית לפני מספר שנים;

4. חיזוק הדמוקרטיה ומאבק בשחיתות: באמצעות בדיקות של מידע ונתונים, כלי בינה מלאכותית מאפשרים למנוע דיסאינפורמציה ומתקפות סייבר ולשפר את השקיפות והתחרות, למשל בתהליכי רכש ציבוריים. אל-חזבי²⁴, לדוגמה, חקר את תופעת ה"טרולינג" ברשתות חברתיות בשל השפעתה הניכרת על המוניטין של מוסדות ציבור והגברת הקיטוב. הנריקו ואחרים²⁵ הציעו להשתמש בכלי בקרה המבוססים על למידת מכונה ועל סטטיסטיקה כדי לגלות אי-עמידה בדרישותיהם של חוזים ציבוריים, שעלולה להיות לה השפעה חמורה על מוסדות;

5. שיפור הבטיחות וביטחון הציבור: בינה מלאכותית כבר מסייעת במניעת פשיעה ומשמשת את מערכת המשפט הפלילי לעיבוד מהיר וניתוח מדויק יותר של מעשים פליליים שיכול לצמצם הליכים שיפוטניים, לרבות מניעה או חיזוי של מתקפות טרור או מעשים פליליים אחרים. בינה מלאכותית גם משולבת בפעילות ביטחונית צבאית או לאומית במדינות רבות.

כאשר טכנולוגיה דיגיטלית מתוכננת ומנוהלת כראוי, היא יכולה לסייע בהטמעת דרכי עבודה המחזקות אחריות, שוויון וניטרליות במינהל הציבורי. ממשלות יכולות לזהות את צורכי השירות, להקצות משאבים בצורה שוויונית ולחזק את הלגיטימיות הדמוקרטית באמצעות צמצום הטיות ושחיתות בעזרת שימוש בנתונים בזמן אמת וניתוחם

בהתבסס על בינה מלאכותית²⁶.

כדי להתגבר על הקשיים הנגרמים ממערכות קודמות (לגאסי), אינרציה ביוקרטיית ותשתיות בלתי מתאימות, לאוקדיו²⁷ מציע להעדיף שילוב הדרגתי של השימוש בבינה מלאכותית בהליכי הביקורת הציבורית. כדי לקדם תרבות של למידה מתמשכת והסתגלות, הוא מדגיש את הצורך בהכשרה יסודית של המבקרים - הן ליישום הטכני של השימוש בכלי בינה מלאכותית והן לפרשנות של התובנות המופקות באמצעות כלים אלה במסגרת תקני ביקורת מקובלים. המתודולוגיה של שילוב הדרגתי משפרת במובהק את רמת הדיוק של הערכות הסיכונים, קבלת ההחלטות ותוצאות הביקורת ככלל, דבר שיוביל לניהול יעיל יותר של משאבי ציבור.

חלק ממאפיינים אלה קשורים ישירות לתפקידים שמבצעים מוסדות ביקורת עליונים, כמו הפחתת חוסר יעילות, שיפור הניהול הציבורי, מאבק בשחיתות והונאה או הגשת הצעות לשיפור התכנון וההשפעה של מדיניות ציבורית באמצעות הערכה וביקורת ביצועים. על כן, בפרק הבא נתמקד בכמה כלי בינה מלאכותית אשר שימושיים לא רק עבור המגזר הציבורי אלא גם עבור הבקרה החיצונית על ניהול גופי ציבור.

ואולם, היתרונות הפוטנציאליים של מערכות בינה מלאכותית עבור המגזר הציבורי כרוכים בשורה של סיכונים בלתי ניתנים לעקיפה, בעיקר בהיבטים הבאים:

- פרטיות, סודיות ואבטחה: איסוף ועיבוד של כמויות נתונים גדולות במגזר הציבורי עלולים ליצור אתגרים משמעותיים של פרטיות ואבטחה. במערכות בינה מלאכותית עלולים להתקיים פגמים וחולשות שיש לצפות מראש כדי למנוע גישה בלתי מורשית, כמו מתקפות המנצלות את יכולתן של המערכות ללמוד או לפעול על בסיס הידע שרכשו. חיוני לוודא שנתוני האזרחים מוגנים ושלא נעשה בהם שימוש לרעה. חיוני גם ליידע את האזרחים על זכויותיהם, על התקנות הרלוונטיות ועל הדרך שבה יוכלו להגיש תלונות אם יראו בכך צורך^{26,27}.

- שקיפות: לעיתים קרובות יש קושי לפרש את התוצרים של מערכות בינה מלאכותית כמו רשתות עצביות עמוקות. שקיפות אלגוריתמית היא האלמנט שבאמצעותו אזרחים יכולים לדעת כיצד מערכות החלטה אוטונומיות מחליטות החלטות המשפיעות על חייהם²⁸. חלק ניכר מעיבוד המידע, אחסונו והשימוש בו מבוצע על ידי אלגוריתמים, ובדרך שאינה שקופה - בתוך "קופסה שחורה" שלמעשה לא ניתן לחקור אותה ותוכנה אינו ידוע אפילו למתכנתים שלה²⁹. מצב זה מעורר חששות בעניין היעדר שקיפות

בהחלטות המדינה, שפעמים רבות נושאות השלכות על חייהם של פרטים או קבוצות, ומקשה להסביר ולהבין כיצד מתקבלות החלטות מסוימות. כפי שציין בריהיל וחוקרים נוספים³⁰ במחקרם, כאשר האלגוריתמים מורכבים מדי אפשר לחזק את האפשרות להסביר אותם בעזרת מנגנוני עקיבות וביקורת, לצד גילוי נאות לגבי היקפם;

- אפליה אלגוריתמית, נגישות ושוויוניות: אלגוריתמים של בינה מלאכותית עלולים להנציח הטיות ואפליה אם הם מאומנים באמצעות נתונים היסטוריים שגויים או מוטים המשקפים הטיות או דעות קדומות של האנשים שאוספים אותם. הדבר עלול להשפיע על נגישותם של משאבים או שירותים עבור קבוצות אזרחים שונות לפי מגדר, גזע, גיל או גורמים אחרים, על רמת הפיקוח שאליה חשופות קבוצות אלה ואפילו על יכולתן להילקח בחשבון בסביבה שבה שמור לטכנולוגיות מקום מרכזי³⁰. כך, אלגוריתמים עלולים להעצים עוולות הנובעות מהטיות חברתיות³¹, לחלק משאבים בצורה בלתי שוויונית³² ולחזק את האופי הטכנוקרטי של המינהל הציבורי³³.

ראוי לציין, לדוגמה, כי פחות מאחוז אחד מבוגרי התארים בתחום הבינה המלאכותית בחרו במשרות ממשלתיות על פני משרות באקדמיה או במגזר הפרטי בארצות הברית. פער כשרונות זה מעכב באופן ישיר את הטמעתן של תוכניות מבוססות בינה מלאכותית - מטכניקות ביקורת מתקדמות ועד ניתוח מדיניות מבוסס ראיות. הוא מאלץ סוכנויות כמו רשות מס ההכנסה של ארצות הברית (IRS) להסתמך על מתמחים המגיעים לתקופה קצרה מכוח חוק חילופי עובדים בין גופים ממשלתיים (IPA) לצורך פיתוח היכולות בתחום מדעי הנתונים והבינה המלאכותית, הנדרשות למשל להערכת מסדי נתונים של ביקורות מס²⁸.

כפי שצוין בדוח של CAF²², הערכת ההשפעות הפוטנציאליות של השימוש באלגוריתמים, יחד עם זיהוי הסיכונים הקשורים בו, היא אלמנט חיוני בקביעת הפעולות הנחוצות להבטחת עמידה בכללי האתיקה בהקשר של בינה מלאכותית.

לעניין זה, יש ממשלות שכבר מעודדות או דורשות נקיטת גישות סדורות (good practices) בכל הנוגע לשימוש בנתונים וניתוח של הטכניקות המיושמות. דוגמאות לכך הן מסגרת העבודה למדעי הנתונים ואתיקה (Data Science Ethical Framework) של ממשלת בריטניה או - עבור חברי האיחוד האירופי - התקנה הכללית האירופית להגנת מידע (GDPR) מ-2018 וחוק הבינה המלאכותית (AI Act) שנכנס לתוקף ב-2023, אשר בהיותו החוק הראשון מסוגו בעולם,

יכול לשמש כמדד השוואה גלובלי להסדרת תחום הבינה המלאכותית בתחומי שיפוט אחרים.

ה-AI Act מחזק באופן ניכר את החקיקה בכל הקשור לפיתוח בינה מלאכותית ושימוש בה, והוא מתמקד בהסדרת הבינה המלאכותית כאשר היא מסוגלת לגרום נזק לחברה, וזאת בגישה מבוססת-סיכונים - ככל שהסיכון גבוה יותר התקנים מחמירים יותר. באמצעות תקנים טכניים מחמירים, ה-AI Act עשוי גם לבסס את מעמדה של אירופה כמובילה עולמית בשימוש בבינה מלאכותית מהימנה. ואולם, קובעי מדיניות חייבים לתת מענה אקטיבי לאתגרים ולחוסר האיזון שעומדים מתמודדים ספקי בינה מלאכותית בגבולות האיחוד האירופי³⁴.

אפילו ארגונים רב-צדדיים, כמו ארגון האומות המאוחדות, הבנק העולמי ואונסק"ו העלו חששות לגבי הסיכונים הנובעים ממערכות בינה מלאכותית, וגם הם משקיעים מאמצים ביצירת מסגרות עבודה וקווים מנחים לבינה מלאכותית בטוחה ומהימנה. באותו האופן, גם ה-AI Observatory של מדינות ה-OECD מנתח ודן בסוגיות חשובות בתחום הבינה המלאכותית.

ברמה הלאומית, אנו עדים ליוזמות ציבוריות ופרטיות הממוקדות בחשש הציבורי הגובר וכן בפיקוח על השימוש בבינה מלאכותית במגזר הציבורי ובביקורת בתחום זה, לדוגמה NESTA (סוכנות החדשנות למען החברה בבריטניה) או ETICAS (קרן פרטית העוסקת בביקורת של מערכות בינה מלאכותית הפועלת בספרד).

על רקע זה, ביסוסם של מנגנוני ממשל פנימי אפקטיביים לבינה מלאכותית במגזר הציבורי הוא רלוונטי במיוחד - זאת לאור אופייה המתפתח של הבינה המלאכותית, רמות הבשלות השונות של הטכנולוגיה המשמשת במגוון מערכות בינה מלאכותית ורמת אי-הוודאות הגבוהה בעניין מגבלות השימוש במערכות אלה וההשפעות והתוצאות האמיתיות שלהן³⁵. התהליך לביסוס מסגרת ממשל יכול לכלול אלמנטים כגון מדיניות ותקנות אסדרה המגדירות את העקרונות ואת התקנים לפיתוח, ליישומן ושימוש בהן; הליכים ומנגנונים המסייעים להבטחת יישום אפקטיבי; ומוסדות המקדמים פיתוח והטמעה של מדיניות ותקנות אסדרה, כמו גם מבני ממשל שיתופיים המבוססים על השתתפותם של מגוון בעלי עניין בעיצובן של מערכות בינה מלאכותית. המטרה היא להרחיב את האחריות, ליצור למידה מעמיקה כדי לגבש פעולות מניעה נגד הטיות ועוולות ולפתח מנגנוני מעקב אחר המערכות³⁵.

נוסף על כך, השימוש בבינה מלאכותית עתיד להוביל לשינויים מבניים בכוח האדם במגזר הציבורי. מחד גיסא, אוטומציה המבוססת על בינה מלאכותית יכולה להחליף משרות מסוימות, מה שמעורר חששות בנוגע לאבטלה ולצורך לרענן ולעדכן את מיומנותיהם של עובדי ציבור ולהכשיר אותם לתפקידים חדשים.

מאידך גיסא, מכיוון שכמה מהחסמים המבניים המקשים על אימוץ בינה מלאכותית במגזר הציבורי טמונים בכישורים הנדרשים לשם כך, יש צורך לעדכן ולהשביח אותם באופן מתמשך. לפי הנציבות האירופית³⁶, ניתן לחלק את הכישורים הללו לשלוש קטגוריות, כמוצג בתרשים 1.

תרשים 1. הכישורים הנדרשים להתמודדות עם בינה מלאכותית במגזר הציבורי. מקור: הנציבות האירופית, 2024, מרכז המחקר המשותף [36].



שאלות אלה מיועדות לשמש מסגרת עבודה צופה פני עתיד עבור מוסדות ביקורת עליונים המבקשים להיערך לעתיד שבו הביקורת של מערכות בינה מלאכותית תהפוך בוודאות לחלק בלתי נפרד ממטלותיהם. אנו מכירים בכך שהמודעות ורמות המיומנות בכל מוסד משתנות באופן ניכר ממדינה למדינה וממגזר למגזר. לכן, השימוש בחקרי מקרה של מוסדות ביקורת עליונים שונים עשוי לסייע בעיצוב האסטרטגיה העתידית של מספר מוסדות ביקורת.

פרק שלישי זה מבקש אפוא להתמקד בשיקולים אלה, ולבחון ולחשוף מספר מקרים כדוגמאות המשקפות התמודדות עם האתגרים המדוברים במוסדות ביקורת עליונים. במובן זה, המטרה היא להשתמש בחקרי המקרה כדוגמאות שיש בהן ערך למוסדות ביקורת עליונים ולקהילות ביקורת אחרות. הם עשויים לשמש גם חוקרים אחרים ופועלים לדברי החוקרת מורגן³⁸ כ-"crystallizers" - תכנים המסייעים לגבש את הידע בנושא.

3.2 אסטרטגיות מוסדיות

הטמעתם של כלי בינה מלאכותית במוסדות ביקורת עליונים איננה פיתוח טכני גרידא, אלא חלק מהסתגלות מוסדית רחבה יותר לטרנספורמציה הדיגיטלית המתחוללת בגופים המבוקרים ובמגזר הציבורי על כל חלקיו. חלק זה של המאמר חוקר כיצד הטמעתם של כלים אלה מתיישבת עם תפקידם המתפתח של מוסדות ביקורת עליונים ותומכת ביכולתם לבקר סביבות שהולכות ונעשות מורכבות יותר ומבוססות נתונים.

כמו בתחומים אחרים, שילוב בינה מלאכותית בעבודת הביקורת מייצר הזדמנויות גדולות לשיפור האפקטיביות של הבקרה הפנימית והחיצונית, כמו גם לשיפור הפרודוקטיביות של העבודה המבוצעת. בפועל, זהו צעד נוסף בשילובה של טכנולוגיה מתקדמת כמכשיר חיוני עבור המבקר. אם כי במקרה זה, עם אלמנטים מבדלים מסיימים המוסיפים מורכבות, אף שהם יספקו קרוב לוודאי תשואות גבוהות יותר בטווח הבינוני והארוך.

למעשה, עד כה, טכנולוגיות חדשות סיפקו למבקרים כלים חיוניים להגברת האיכות והפרודוקטיביות של עבודתם, החל מגיליון אלקטרוני פשוט או מעבד תמלילים ועד ליישומים אלקטרוניים מורכבים לניהול ביקורת. בשנים האחרונות, השימוש בפיתוח ביג דאטה התפשט לכל שלבי מחזור הביקורת, והוא מאפשר עבודה רחבה יותר ועמוקה יותר בכל אחד מן השלבים.

כך, במהלך שלב התכנון, ניתוח נתונים מאפשר איתור תחומי סיכון, בחירת נושאים רלוונטיים לביקורת, זיהוי מאגרי נתונים מתאימים וקביעה אם הם עומדים בדרישות או מצריכים טיפול כדי להשתמש בהם בפיתוח הביקורת.

כל אלה שייכים למסגרת של אסטרטגיית פיתוח משאבי אנוש לטווח הארוך בשירות הציבורי, המביאה בחשבון את סיווג המשרות, תוכניות ההכשרה וסוגי הפרופילים והמיומנויות המבוקשים³⁷ בתרחיש חדש זה.

מודל כשירויות זה יכול להועיל מאוד למוסדות ביקורת עליונים. הוא מצביע על סוגי הכישורים הדרושים כדי להתמודד עם האתגרים הקשורים לבינה מלאכותית, העתידים להתגבר ככל שהמוסדות ירחיבו את השימוש בה (לדוגמה, אוריינות ומיומנויות טכניות, תכנון אסטרטגי ויישום נורמות ראיות).

מכל מקום, מוסדות ביקורת עליונים יצטרכו לבדוק ולוודא באמצעות עבודת הביקורת שלהם כי שילוב בינה מלאכותית במגזר הציבורי מתרחש בתוך מסגרת ממשל נאותה, החותרת למזער סיכונים העלולים להתעורר, כמו הופעת הטיית, חוסר צדק וחוסר שוויוניות, חוסר ודאות לגבי השפעות הממשיות של הטמעת הבינה המלאכותית ועוד. פעילות זו תעסוק גם בבקרה על היבטים נוספים שהוזכרו קודם, כמו מדיניות ותקנות אסדרה, מוסדות, נהלים ומנגנונים הקשורים לבינה מלאכותית ולשימוש בה במגזר הציבורי.

אין ספק אפוא כי מוסדות ביקורת עליונים ברחבי העולם צפויים לעמוד בשנים בקרובות בפני אתגר גדול ועליהם להיות ערוכים לכך.

3. בקרה חיצונית על ביצועי המגזר הציבורי לאור השימוש בבינה מלאכותית

3.1 אתגרים מרכזיים באימוץ בינה מלאכותית

מוסדות ביקורת עליונים - המוסדות העצמאיים האמונים על הבקרה החיצונית של הניהול והפעילות הפיננסית של גופים ציבוריים אינם יכולים לנתק את עצמם מההתקדמות המהירה של הבינה המלאכותית בתחומים רבים, ובפרט במגזר הציבורי. חשוב להכיר בשינויים המבניים שיחולו בעקבות הבינה המלאכותית בניהול הגופים הציבוריים המבוקרים. כמו כן, יש לבחון מהי האסטרטגיה שעל מוסדות הביקורת לגבש עבור השנים הקרובות במטרה לנצל את התרומה שהבינה המלאכותית יכולה להציע לשיפור האפקטיביות והיעילות במשימות הביקורת ובתחומים אחרים של הארגון ולהתמודד עם האתגרים אשר ללא ספק יתעוררו בעתיד הלא רחוק.

אסטרטגיה כזו צריכה לענות לכל הפחות על שתי שאלות:

- אילו מערכות בינה מלאכותית יכולות לשמש ככלים לביצוע העבודה בארגון בצורה אפקטיבית ויעילה יותר?
- כיצד על הארגון להיערך לביצוע ביקורת על מערכות בינה מלאכותית המשמשות לניהול שירותים לציבור?

להרחיב את היקף הביקורת, לשפר את איכות הניתוח ולהימנע מבעיות בקרה פנימית". אותה סקירה בוחנת שני תנאים נוספים הרלוונטיים לשימוש בטכניקות של ניתוח ביג דאטה במוסדות ביקורת עליונים:

- הסדרים מוסדיים, כמו תקנות אסדרה או החלטות במינהל הציבורי, הכרחיים לקידום שיתוף נתונים, להותרת הנתונים פתוחים ולשיפור הערך המופק מהשימוש בהם;
- בניית פלטפורמת ביקורת במוסד הביקורת המשלבת את כלל התהליך של איסוף הנתונים, הכנתם, אחסונם, ניתוחם והצגתם תאפשר ניהול משולב של המבקרים, נוהלי הביקורת והאמצעים הטכניים במערכת יחידה.

אין ספק שהיכולת של הבינה המלאכותית לעבד ביעילות כמויות גדולות של נתונים לא רק הגבירה את ערך הנתונים הנאספים במהלך הביקורת, אלא גם הורידה את העלות ואת זמן העיבוד של כמויות הנתונים הגדולות הללו³⁹.

למעשה, מערכות הבינה המלאכותית הנפוצות ביותר בתחום הביקורת מבוססות על טכניקות של למידת מכונה (ML) - הבניה כאמור על אלגוריתמים לומדים שמטרתם להשיג תוצאה התלויה במשתני הקלט של המודלים (נתונים), במיוחד למידת מכונה מונחת (supervised ML). לוח 1 להלן המעובד על בסיס המאמר של ריוורה⁴⁰, מציג כמה מכלי למידת המכונה הנפוצים ביותר. זהו למעשה סיכום השוואתי של כלי למידת המכונה העיקריים הרלוונטיים לפונקציות ביקורת חיצונית ושל יישומיהם הפוטנציאליים במוסדות ביקורת עליונים, עם התאמות המדגימות את הפוטנציאל של כל אחד מן הכלים באופן ספציפי בפונקציות הבקרה החיצונית של מוסדות ביקורת עליונים. עבור כל יכולת אלגוריתמית, הטבלה מציגת תרחישי ביקורת מהעולם האמיתי, כך שהיא מתייחסת ישירות לדוגמאות המעשיות המוצגות בפרק זה ומראות כיצד מוסדות ביקורת עליונים משתמשים בטכנולוגיות הללו כדי לזהות תחומי סיכון, למכן (להפוך לאוטומטיות) מטלות חזרתיות (רפטיטיביות) ולשפר את האפקטיביות והיעילות של הבקורות שלהם.

מכל מקום, מערכות למידת מכונה זקוקות לאימון באמצעות כמויות גדולות של נתונים מובנים ובלתי מובנים, העונים על שורה של דרישות מוגדרות מראש, אשר באמצעותן יש למצוא דפוסים שהוגדרו מראש בידי המבקר. משום כך, הטמעתן של מערכות אלה בשטח חייבת להתאים לכל מקרה ספציפי, לפי המטרה המבוקשת והמידע הזמין. בתמורה, המערכות מאפשרות לבצע מטלות חזרתיות, כך שהן מחליפות התערבות אנושית - אם כי הן מצריכות (לפחות בתחילת השימוש) ניטור מסוים של התוצאות המופקות.

בהמשך, בשלב היישום, נתונים יכולים לשמש לבניית מודלים המאפשרים להשיג ראיות או לגלות טעויות, כמו גם להצליב בין מיליוני נקודות נתונים ולקבל תוצאות ומסקנות ביקורת, וזאת ללא צורך בשימוש במדגמים, כל עוד זמינות הנתונים מסוג זה מאפשרת זאת.

בתחום הבקרה החיצונית, הצהרת מוסקבה של ארגון הביקורת העולמי INTOSAI שניסחה את המסקנות העיקריות מוועידת INCOSAI העשרים ושלוש שהתקיימה ב-2019, מדגישה את מחויבותם של מוסדות ביקורת עליונים להגיב בצורה אפקטיבית להזדמנויות הנולדות מתוך התפתחויות טכנולוגיות. לעניין זה, מתוך ההכרה כי ניתוח נתונים הוא חידוש הכרחי במוסד הביקורת, ההצהרה מחייבת מוסדות "לקדם את העיקרון של זמינות ופתיחות נתונים, קוד מקור ואלגוריתמים" וכן "לחתור לשימוש משופר בניתוח נתונים בביקורת, לרבות גיבוש אסטרטגיות לניתוח נתונים והנהגתן של טכניקות חדשות בפרקטיקה של הביקורת הציבורית". נוסף על כך, במטרה לחזק את השפעתם של מוסדות ביקורת עליונים על החברה, ההצהרה מעודדת הכשרה של מבקרים עתידיים כדי ש"יוכלו להשתמש בניתוח נתונים, בכלי בינה מלאכותית ובשיטות איכותניות מתקדמות לחיזוק החדשנות...".

אם כן, זמינות הנתונים ואיכותם הם אלמנטים אסטרטגיים בכל מוסד בקרה, מאחר שהם יכולים לשמש בסיס לביצוע ניתוח סיכונים ולגילוי מקרים של מרמה או אי-סדרים בניהול. הדבר מאפשר למוסדות להיות אפקטיביים ויעילים יותר בביקורת, ויכול גם לסייע בקידום אחריותיות ושקיפות במינהל הציבורי, לצד עמידה טובה יותר בעקרונות של ניהול הולם.

נוסף על ההשפעה הישירה של השימוש הישיר בנתונים - חיסכון זמן ומשאבים והימנעות מטעויות - עשויה להיות חשיבות גם לחילופי נתונים בין מוסדות ביקורת עליונים ומוסדות אחרים, כך שיוכלו לקדם שיתופי פעולה שיעודדו חדשנות ותכנון של פתרונות וכלים משותפים.

בעת הזאת, קיומם של נתונים איכותיים הוא הכרחי כדי לאפשר שימוש בכלי בינה מלאכותית כחלק מעבודת הביקורת, מגמה שתגדל בקצב ניכר בשנים הקרובות. מודלים של בינה מלאכותית משתמשים במאגרי נתונים גדולים - מובנים ובלתי מובנים - כדי להפיק תוצאות נכונות ומהימנות. אולם אם איכות הנתונים אינה גבוהה, תוצאות המודלים עלולות להיות שגויות או מוטות. לפיכך, חשוב שבמאגרי הנתונים יתקיימו המאפיינים הנדרשים.

בהקשר זה, סקירת ביקורות ביג דאטה (Overview of Big Data Audits) שביצעו מוסדות ביקורת עליונים בשנים 2016 - 2021³⁹ ושהכינה קבוצת העבודה של INTOSAI לנושא ביג דאטה קובעת כי "האיכות והאבטחה של נתוני ביקורת חיוניות להבטחת ביצוע תקין של ביקורת ביג דאטה. שיפור האיכות והאבטחה של נתונים יכול לסייע

סוג כלי	שימוש כוטנציאלי
אלגוריתמי אשכול (clustering) של נקודות נתונים דומות	בבקרה חיצונית, אלגוריתמים אלה יכולים להועיל, לדוגמה, לסיווג הוצאות לפי משרד ממשלתי או לזיהוי קבוצות של תוכניות תקציביות או פרויקטים דומים, וכך להקל על ההשוואה וההערכה של פריטים אלה
אלגוריתמים לגילוי חריגות או סטיות מהתקן	שימוש בכלי זה לבקרה מאפשר לגלות חריגות תקציביות או לתעדף נושאי ביקורת בהתאם לתחומים שבהם התוצאות חורגות מהדפוסים הצפויים
רשתות עצביות מלאכותיות	אלגוריתמים אלה מאפשרים לבצע מטלות שונות, לרבות זיהוי תמונה וזיהוי קולי ועיבוד שפה טבעית (NLP - natural language processing). בתחום הבקרה החיצונית ניתן להשתמש בהם לעיבוד ולניתוח כמויות גדולות של מידע בלתי מובנה, כמו טקסטים ותמונות, במטרה לחלץ מהם נתיבי ביקורת או מסקנות רלוונטיות. לדוגמה, ניתן ליצור מודלים של ניבוי מתוך מידע הלקוח מביקורות קודמות כדי לגלות תשלומי סיוע שהוענקו ללא אישור, הוצאות חריגות או תמחור יתר
עצי החלטה	במסגרת בקרה חיצונית, כאשר מסווגים נקודות נתונים על בסיס כללי החלטה שהוגדרו מראש, ניתן להשתמש במערכת לסיווג עסקאות כהונאה או כעסקה כשרה, לסווג ספקים כבעלי סיכון גבוה או נמוך או לנבא את הסבירות להונאה בתחום מסוים.
K השכנים הקרובים ביותר	אלגוריתמים אלה נמנים על האלגוריתמים הפשוטים ביותר והקלים ביותר ליישום, והם משמשים לסיווג נקודות מסוימות (שכנים) החל מנקודות אחרות. k הוא מספר השכנים שנבדקים. כלים אלה משמשים ביישומי זיהוי תמונות וסרטונים, ניתוחי בורסה, זיהוי דפוסים, גילוי פולשים ובנייה של אלגוריתמים מורכבים יותר

של אותן משימות תאפשר ביצוע יותר בדיקות בכחות זמן, וכך תושג התייעלות. RPA מסוגל לזהות חוסר עקביות או ממצאים חריגים במידע המוגש במהלך הביקורת.

בשלב זה יש מקום לנתח כיצד מוסדות ביקורת עליונים מסוימים כבר משלבים כלי בינה מלאכותית בפעילויות הבקרה שלהם. ניסיונם ישמש נקודת ייחוס עבור מוסדות אחרים, שיידרשו אף הם לאמץ טכנולוגיות אלה בעתיד הקרוב, אם אינם משתמשים בהן כבר עתה.

בדרך זו, באמצעות זיהוי חריגות ומגמות שמובילות לגילוי תחומי סיכון, הבינה המלאכותית מייעלת את עבודתם של צוותי ביקורת, שיכולים להיות פרודוקטיביים יותר ולהקדיש זמן רב יותר למטלות שבהן הניסיון והידע שלהם תורמים ערך גבוה יותר לתוצר הסופי. הצורך במטלות חזרתיות מתבטל וטווח הטעות בתוצאות קטן.

מעניין גם לבחון שימוש באוטומציה רובוטית של תהליכים (RPA) למשימות ביקורת הנוטות לחזרתיות. האוטומציה

מסקנות הביקורת. מוסד זה משתמש גם באוטומציה רובוטית של תהליכים (RPA) כדי לזהות חוסר עקביות או ממצאים חריגים שהמבקרים האנושיים יכולים לטפל בהם בהמשך.

המקרים שנותחו הם רק דוגמאות ספורות לשימוש של מוסדות ביקורת עליונים בכלי בינה מלאכותית למיכון מטלות, לגילוי תחומי סיכון ולפעולות נוספות. מרבית המקרים שהוצגו חושפים כי כלי הבינה המלאכותית הנפוצים ביותר בשימוש הם בעלי זיכרון מוגבל ושייכים לסוגים הצרים של בינה מלאכותית. אלה מבוססים על למידת מכונה, בעיקר לצורך מיכון בדיקות, ניתוחי מידע וחילוץ אוטומטי של מידע ממקורות שונים. הדוגמאות שהובאו כאן גם חושפות הטרגויות ניכרת במוכנות של מוסדות ביקורת עליונים לביקורת באמצעות מערכות בינה מלאכותית - המשקפת הבדלים ביכולות מוסדיות, משפטיות וטכניות. בשעה שמוסדות מסוימים, כמו אלה של ברזיל, הפיליפינים או ארצות הברית, נוקטים צעדים יזומים לשילוב כלי בינה מלאכותית בתהליכי הביקורת שלהם, אחרים נמצאים עדיין בשלבי בחינה ומיפוי.

במובן זה, חקירת המצב הנוכחי בתחום זה חיונית לאור התפתחותן המהירה של טכנולוגיות חדשות, וזאת כדי לנסות ללמוד מניסיונם של מוסדות בקרה אחרים ולהאיץ את תהליך האימוץ והשימוש בבינה מלאכותית במוסדות ביקורת עליונים. בכל הדוגמאות, האתגרים הנפוצים כלל פועלים במיומנות הטכנית, מגבלות בממשל נתונים (data governance) והיעדר מסגרות ביקורת תקינות למערכות האלגוריתמיות. ממצאים אלה מצביעים על הצורך במאמצים בין-לאומיים מתואמים, במנגנוני שיתוף ידע ובבניית יכולת ממוקדת כדי לגשר על הפערים.

3.3 דוגמאות לשימושים בבינה מלאכותית במוסדות ביקורת עליונים

ככל שהפיתוח וההטמעה של מערכות בינה מלאכותית בגופים במינהל הציבורי גוברים בקצב מסחרר - בזכות הפוטנציאל הטמון בהם לשיפור השירותים לציבור ולהפחתת עלויות ובזכות יתרונות נוספים - מתעוררים גם אתגרים וסיכונים חדשים, כמו הטיות המובילות לאפליה או ליחס בלתי שוויוני, הצורך להבטיח את הגנת המידע והפרטיות, קבלת החלטות שגויה המבוססת על תהליכים אוטומטיים ועוד. הבינה המלאכותית מציעה הזדמנויות חסרות תקדים לנתח כמויות אדירות של נתונים במהירות ובדייקנות, אך היא גם מכניסה מורכבויות חדשות של ממשל נתונים, הטיות ואיתקה.

בשל כך, מוסדות ביקורת עליונים חייבים להתאים את מבנה הארגון לביקורת של יישומים מבוססי אלגוריתמים כחלק מביקורת תפעוליות או ביקורת ציט, וזאת כדי להמשיך ולשמש כמגינים אפקטיביים של אחריות ציבורית.

מספר מוסדות ביקורת עליונים מתחילים לעבוד באופן אינטנסיבי יותר עם כמויות גדולות של נתונים, לעיתים באמצעות מרכז ניהול נתונים או מחלקה ספציפית בארגון המופקדת על ניתוח נתונים וטכניקות מתאימות.

כך לדוגמה, מוסד ביקורת עליון בארצות הברית מנהל מאז 2019 מעבדת חדשות כדי לחקור ולהתנסות בטכניקות של מדעי הנתונים וטכנולוגיות חדשות (<https://www.gao.gov>). בדומה לכך, מוסד ביקורת עליון בהודו מפעיל מרכז ניהול וניתוח נתונים (CDMA) עבור כל הפעילויות הקשורות לניתוח נתונים, מספק הנחיות לניתוח נתונים במשימות ביקורת נבחרות, מציע את ההכשרה הדרושה לעובדים ומפתח אסטרטגיה לזיהוי היקפו העתידי של ניתוח הנתונים במוסד. המרכז אוסף נתונים ממערכות טכנולוגיות מידע ותקשורת (ICT) מרחבי המדינה כדי לנתח אותם ולזהות תחומי סיכון פוטנציאליים, וזהו מידע עזר לביקורת עתידית (<https://cag.gov.in/en/page-cdma>).

יש מוסדות ביקורת עליונים שכבר עובדים עם רובוטים, כמו מוסד הביקורת העליון בברזיל, אשר עובד מאז 2017 עם רובוטית בשם "אליס" שמנתחת מכרזים ציבוריים ועסקאות רכש כדי להאיץ את איסוף הנתונים לצורך זיהוי אי-סדרים במכרזים ציבוריים, כך שהיא מבצעת את העבודה המייגעת והחזרתית. לרשות מוסד זה עומדים גם כלים אחרים, כמו Labcor - המתמקד במודיעין ובפעולות נגד שחיתות ו-SAIO - כלי שמנתח אוטומטית תקציבים של עבודות ציבוריות. יוזמה מעניינת נוספת נקראת Marina (מפת סיכונים ברכש הציבורי), והיא מסייעת במניעת שחיתות בחוזים ציבוריים באמצעות פיקוח על הסיכון הקשור אליהם. כלי זה מחפש סימנים המלמדים על ליקויים או חולשות במכרזים או בחברות הזוכות.

בינה מלאכותית משמשת גם במיכון של מטלות ביקורת מסוימות במוסד הביקורת העליון בפיליפינים, שפיתח את פלטפורמת הבינה המלאכותית MIKA-EL המסוגלת לזהות אוטומטית עסקאות חריגות או לא תקינות. היא בודקת מיליוני פעולות על פי נתונים הנאספים מגופים מבוקרים ומבצעת בדיקה צולבת מול נתונים ברמות מינהליות אחרות (<https://www.coa.gov.ph/coa-explores-the-use-of-ai-to-detect-statistical-anomalies/>). מוסד הביקורת העליון של ספרד מיכן זיהוי חשבוניות הקשורות לתהליכים אלקטרוניים כדי לשלב את התוצאות במאגרי הנתונים של המוסד. הדבר מאפשר אוטומציה והאצה משמעותית של ניתוח ההצבעה בבחירות והבקרה על ספירת הקולות.

לסיום, מוסד הביקורת העליון של נפאל משתמש בטכניקות בינה מלאכותית לביצוע מגוון מטלות ביקורת⁴². לדוגמה, הוא משתמש בזיהוי תווים אופטי (רשתות עצביות מלאכותיות) כדי למכן חילוץ מידע קבוע מסוים על פי שדות נתונים במסמכים בעלי פורמטים שונים וישתמש באלגוריתמים כדי לבדוק התאמות בנתוני גביית המיסים כדי לסייע בהפקת

- הצורך בהבנה נאותה של העקרונות הכלליים של מודלי בינה מלאכותית.
 - הצורך בהבנה של שפות התכנות הנפוצות ומימושי המודלים וכן הצורך במיומנות בשימוש בכלי התוכנה המתאימים.
 - הצורך בתשתית מחשבים לתמיכה בלמידת מכונה בעלת עוצמת מחשוב גבוהה, הכרוכה על פי רוב בפתרונות מבוססי ענן.
 - הצורך בהיכרות בסיסית עם שירותי ענן כדי לבצע את עבודת הביקורת כראוי.
- לצד צרכים אלה, העלולים להציב חסמים פוטנציאליים בעת ביצוע ביקורת הבודקת מודלי למידת מכונה, ניתן להצביע על עוד קשיים בתחום זה:
- ידע מוקדם מועט: היישום של בינה מלאכותית במינהל הציבורי הוא חדש יחסית ואינו נפוץ במיוחד, כך שטרם נצבר ניסיון רב הן בניהול והן בביקורת.
 - מיומנויות נמוכות בקרב מבקרים: מאחר שמדובר בתחום ידע חדש ומורכב, קיים קושי בהרכבת צוות בעל הכשרה מספקת כדי להתמודד עם ביקורת מסוג זה, כך שיש להעמיד עניין זה במקום גבוה בסדר העדיפויות של מוסד הביקורת. ביקורת מסוג זה ממילא קשורה באופן הדוק לביקורת הבודקת מערכות מידע, כך שנוכחותם של צוותים מומחים בתחום זה תקל על המעבר.
 - מיעוט מדריכים וספרי הוראות: עצם החידוש בביקורת המתמקדת במודלים מבוססי בינה מלאכותית יוצר קושי במציאת מדריכים או ספרי הוראות שמתאימים לביקורת מסוג זה.
- לאחרונה פרסמו קוסיאמה וחוקרים נוספים⁴⁴ מאמר מחקר בנושא ביקורת של אלגוריתמים, המתמקד בסיכונים המשפטיים, האתיים והטכנולוגיים הטמונים במערכות בינה מלאכותית. המחברים מזהים חמישה משתני סיכון המחייבים הערכה כאשר נבדקים בביקורת אלגוריתמים: הטיה, יעילות, עמידות (robustness), פרטיות והסבריות (explainability).
- למרות האמור, אין ספק שחשוב כי תחום ידע זה לא יישאר זר לפונקציית הבקרה והפיקוח, ולכן מוסדות ביקורת עליונים מסוימים כבר עושים את צעדיהם הראשונים להטמעתו, אם כי לא בכל המקרים באמצעות ביקורות.
- בחלק מהמוסדות, הדוחות הראשונים שהתפרסמו מתמקדים באסטרטגיה או בעבודתה של הממשלה לפיתוח מערכות בינה מלאכותית. כך למשל בית המשפט האירופי לביקורת (European Court of Auditors) ומוסד ביקורת המדינה של בריטניה, כפי שיתואר להלן.

בהקשר זה, גרדה רוקה⁴³ מדגיש כי על ביקורות מסוג זה לאמץ גישה פרגמטית במיוחד המתמקדת בפריטים חיוניים. הביקורת צריכה לנתח ולהעריך את האלגוריתמים כדי לבדוק את דרך פעולתם ולגלות אם הם ממלאים את מטרותיהם המוצהרות או מייצרים תוצאות מוטות וכגיעויות חברתיות חדשות שאינן מטופלות בידי המסגרת המשפטית והרגולטורית הקיימת.

עם זאת, הביקורות חייבות גם להתייחס להיבטים ספציפיים בתרומתן של מערכות בינה מלאכותית לשיפור היעילות והאפקטיביות של הניהול הציבורי, למניעת שחיתות ושימוש לרעה בכספי ציבור ולהגברת השקיפות והאחריותיות במגזר הציבורי.

מכל מקום, ביקורת של האלגוריתמים העומדים מאחורי הכלים שבשימוש במגזר הציבורי היא תחום שטרם נחקר בידי מוסדות ביקורת עליונים, בשל גורמים שונים המעכבים את הטמעתם המעשית של אותם כלים בתחום הבקרה החיצונית. כדי לקדם משימה זו, פורסמה הגרסה הראשונה של המדריך לביקורת אלגוריתמים של למידת מכונה (Auditing Machine Learning Algorithms) בשנת 2020 - סקירה טכנית למבקרי ציבור (A White Paper for Public Auditors) (<https://www.auditingalgorithms.net>). את הסקירה ערכו מוסדות ביקורת המדינה של פינלנד, גרמניה, הולנד, נורווגיה ובריטניה, ומטרתה "לסייע למוסדות ביקורת עליונים ולמבקרים בביקורת אלגוריתמים של למידת מכונה המשמשים גופים ציבוריים". המדריך מיועד במפורש למבקרים בעלי ידע מסוים בשיטות כמותיות, אך אינו מניח כי הם מומחים במודלים של למידת מכונה. הוא מציג סדרת שאלות שיכולות לשמש את המבקרים בעת שהם עורכים ביקורת ובודקים מודלים של למידת מכונה, ויכולות גם לשמש מנהלים במגזר הציבורי ולהנחות אותם במהלך תכנון מודלים כדי להבין אילו היבטים ישפיעו על הביקורת. כמה מהתחומים הבעייתיים המתוארים בסקירה הם האפשרות של התמקדות יתר בביצועים של מודלים של למידת מכונה המביאה להתעלמות מדרישות מסוימות כמו הוגנות או שקיפות; ליקוי בהעברת המסרים על דרישות המודל בין מנהלים ומפתחים הגורם לתוצאות הפוכות המתבטאות ברמת ביצועים נמוכה ועלויות גבוהות; ואימוץ מודל שאינו בר תחזוקה או כזה שאינו יכול לציית לתקנות האסדרה בטווח הבינוני עקב תלות במפתחים חיצוניים.

מנקודת מבטו של המבקר, המדריך מצביע על השלכות מסוימות עבור מבקרים הנובעות מביקורת של יישומי בינה מלאכותית:

במאי 2024, פרסם בית המשפט האירופי לביקורת דוח מיוחד⁴⁵ שהעריך את האפקטיביות של תרומת הנציבות האירופית לפיתוח המערכת האקולוגית של הבינה המלאכותית באיחוד האירופי. לאחר שבחן את הפעולות שנקטה הנציבות לתיאום ויישום בתחום הבינה המלאכותית ואת האימוץ של מסגרת חוקית אחידה בתחום, הגיע בית המשפט למסקנה כי האמצעים שהנהיגה הנציבות האירופית, ואלה שהונהגו ברמה הלאומית בכל מדינה "לא היו מתואמים באפקטיביות בשל מספרם המועט של כלי הממשל הזמינים, הטמעתם החלקית ויעדים שעבר זמנם. זאת ועוד, ההשקעה של האיחוד האירופי בבינה מלאכותית לא עמדה בקצב בהשוואה לשחקנים מובילים בעולם. הטמעת תשתיות ותמיכה כספית בישויות קטנות ובינוניות (SMEs) כדי שאלה יאמצו טכנולוגיות בינה מלאכותית ארכה זמן, ועל כן לא הייתה תוצאות משמעותיות כשהגיע מועד הביקורת. הנציבות האירופית הצליחה להגדיל את כמות תוכניות המחקר בתחום הבינה המלאכותית הממומנות בידי האיחוד האירופי, אך לא עקבה אחר תרומתן להתפתחותה של מערכת אקולוגית של בינה מלאכותית באיחוד האירופי. מאמצי הנציבות להבטיח שתוצאות המחקר ייתרגמו לחדשנות היו אפקטיביים באופן חלקי בלבד".

מבקר המדינה הבריטי הפיק שני דוחות על יישום בינה מלאכותית במינהל הציבורי הבריטי. הראשון, שפורסם ב-2021, התמקד ב"אתגרים בשימוש בנתונים בממשל"⁴⁶ ואסף מסקנות ממספר דוחות שערך בעניין החשיבות של קבלת החלטות מבוססת ראיות בכל רמות הפעילות בממשל והבעיות שמתעוררות כאשר הנתונים אינם נאותים. הדוח הצביע על שלושה תחומים שבהם ממשלת בריטניה צריכה לקבוע תנאים מוקדמים לשימוש מוצלח בנתונים בקבלת החלטות ובמתן שירותים לציבור:

- קביעת אסטרטגיה ברורה ומנהיגות לשיפור השימוש בנתונים;
- קיומה של תשתית לכידה לניהול הנתונים (בייחוד בהיבטים כמו איכות הנתונים או תאימות בין כלים);
- קיומם של תנאים רחבים יותר (כגון הסביבה המשפטית, הכשרה ואבטחה) להגנה על הנתונים ולתמיכה בשיפור השימוש בהם.

מסקנות הדוח הן שממשלת בריטניה לא גילתה מנהיגות אסטרטגית בת קיימא בנושא נתונים והשימוש בהם לאחר שנים של מאמצים וכישלונות, וכי הראיות המוקדמות לשיפור במצב עלולות להפוך לעוד הזדמנות מוחמצת אם לא יגובשו אסטרטגיית נתונים נאותה ומנהיגות ברורה.

בחודש מרץ 2024, פרסם מבקר המדינה הבריטי דוח על "השימוש בבינה מלאכותית בממשלה", שבחן את האפקטיביות של משרדי הממשלה הבריטית

במקסום ההזדמנויות ובהפחתת הסיכונים הנובעים מבינה מלאכותית במתן שירותים לציבור. הדוח בחן את האסטרטגיה של ממשלת בריטניה ואת תרבות הממשל שלה לשימוש בבינה מלאכותית במסגרת שירותים לציבור. הדוח גם בחן כיצד גופים ממשלתיים השתמשו בבינה מלאכותית וכיצד הם הבינו את ההזדמנויות ואת תוכניות הממשלה המרכזית לתמיכה בבדיקות, בפילוטרים ובהרחבת השימוש בבינה המלאכותית וכן את ההתקדמות בטיפול בחסמים לאימוצה. המסקנות מתייחסות לצורך בהטמעה ואימוץ של בינה מלאכותית בקנה מידה רחב בכל חלקי המגזר הציבורי כדי למקסם את ההזדמנויות שיש לה להציע וקובע כי "הפיתוח והפריסה של בינה מלאכותית בגופים ממשלתיים נמצאים בשלב מוקדם ומתקיימת פעילות לפיתוח אסטרטגיות, תוכניות וכללי ממשל"⁴⁷.

אף על פי כן, יש מוסדות ביקורת עליונים שכבר התחילו לבצע ביקורות הבודקות אלגוריתמים ומערכות בינה מלאכותית ספציפיות אשר נמצאים בשימוש בשירות הציבורי.

מוסד הביקורת העליון בהולנד פרסם ב-2021 דוח בשם "הבנת אלגוריתמים" ("Understanding Algorithms"), שבחן אם הממשלה נמנעה מהטיות באמצעות שימוש באלגוריתמים ואם מתנהל מעקב אחר ההשלכות על האזרחים ובתי העסק במדינה המושפעים ממדיניות ציבורית. מסקנות עבודת הביקורת היו כי אף שהממשלה מכירה בחשיבות הפרטיות, היא אינה מביאה בחשבון כראוי היבטים אתיים, ונטען כי אלגוריתמים אינם נטולי סיכון מאחר ששימוש שגוי או מוטא במאגר נתונים עלול להוביל לאפליה. יתרה מכך, האלגוריתמים ששימשו את משרדי הממשלה היו פשוטים יחסית, עוצבו במיוחד למיכון החלטות ולא התבססו על למידת מכונה. עבודת הביקורת גם הובילה למסקנה שהאלגוריתמים היו כרוכים תמיד במעורבות של אנשים בתהליך הלמידה, אם כי הדבר הגביל את התועלת שניתן להפיק משימוש בבינה מלאכותית במגזר הציבורי. המבקרים מדגישים כי חשוב לשים לב לחששות הציבור וספקותיו בעניין האלגוריתמים, כך שחשוב לאפשר לאזרחים להבין את השימוש באלגוריתמים ואת פעולתם. כמו כן, מאחר שהשימוש של המדינה באלגוריתמים יכול להיות תלוי בספקים חיצוניים (זכויות קניין ועיבוד נתונים אישיים), עליה להבטיח אבטחת מידע נאותה כדי למנוע חבלה, ריגול וכשיעה.

המלצות הדוח מצביעות על הצורך בהגדרה ברורה ועקבית של האלגוריתמים המשמשים בממשל ודרישות האיכות עבורם, וכן על הצורך בפרסום והנגשת מידע על האלגוריתמים לציבור. נוסף על כך, המינהל הציבורי הוזמן להשתמש במסגרת הביקורת שפותחה ושימשה את מוסד הביקורת העליון כדי לפתח אלגוריתמים חדשים ולהתייחס להיבטים כמו ממשל ואחריות, סודיות המודל והנתונים, איכות הבקורות הטכנולוגיות ואתיקה.

מוסד ביקורת עליון של ארצות הברית פרסם, בין היתר, שלושה דוחות מעניינים בנוגע לשימוש במודלים של למידת מכונה במינהל הציבורי בארצות הברית:

- "בינה מלאכותית במערכת הבריאות: יתרונות ואתגרים של טכנולוגיות למידת מכונה לאבחון רפואי" (Artificial Intelligence in Health Care: Benefits and Challenges of Machine Learning Technologies for Medical Diagnostics): דוח זה מנתח טכנולוגיות אבחון רפואי המבוססות על למידת מכונה ומשמשות לאבחון חמש מחלות נבחרות וכן עוסק בטכנולוגיות חדשות, באתגרים המשפיעים על הפיתוח והאימוץ שלהן ובאפשרויות בתחום המדיניות שיכולות לסייע להתמודדות עם אתגרים אלה;
- "משרד ההגנה זקוק להנחיה כלל משרדית לצורך הכוונת הרכש" (Department Of Defence Needs Department-Wide Guidance to Inform Acquisitions): דוח זה בוחן מהם גורמי המפתח שמובאים בחשבון בעת רכישת יכולות בינה מלאכותית, לדבריהן של 13 חברות פרטיות שנבדקו, ובדק באיזו מידה קיימים במשרד ההגנה קווים מנחים לרכישת בינה מלאכותית בתחומי המשרד השונים, וכיצד קווים מנחים אלה משקפים, אם בכלל, את גורמי המפתח שעליהם הצביעו החברות מהמגזר הפרטי;
- "בינה מלאכותית: סוכנויות החלו בהטמעה אך עליהן להשלים דרישות מרכזיות" (Artificial Intelligence: Agencies Have Begun Implementation but Need to Complete Key Requirements): דוח זה מנתח את ההטמעה של בינה מלאכותית בסוכנויות פדרליות מרכזיות, בהתמקד בשימושים נוכחיים וצפויים של בינה מלאכותית שעליהם דיווחו הסוכנויות, דרגת הדיוק של דיווחים בנושא בינה מלאכותית והיותם שלמים ודרגת הציות לכללי מדיניות מסוימים וקווים מנחים פדרליים בנושא בינה מלאכותית.

מוסד ביקורת עליון של ארצות הברית פרסם גם סוגים אחרים של מסמכים הקשורים לנושא זה, ללא ספק כתוצאה מיחסו לביקורת של בינה מלאכותית כתחום מועדף, המצריך רכישת ניסיון וידע הן בתוך מוסד הביקורת העליון והן מחוצה לו.

לפיכך, כבר ננקטים צעדים ראשונים בעריכת ביקורת של מערכות בינה מלאכותית ויישומן בשימושים שונים במגזר הציבורי. מאחר שהדרך בתחום זה עוד ארוכה, דומה כי שיתוף ידע וניסיון בין מוסדות ביקורת עליונים הוא אלמנט חיוני להתקדמות בקצב מספק.

בעקבות כל שנאמר עד כה, נראה כי אין מנוס מהמסקנה שמוסדות ביקורת עליונים ייתקלו בהכרח באתגרים מסוימים בטווח הזמן הבינוני עקב השימוש במערכות בינה מלאכותית במוסדות בקרה ובגופי המגזר הציבורי שעליהם הם מפקחים. בהקשר זה, לפי סקר שביצע בית המשפט האירופי לביקורת⁴⁸, אלה הם האתגרים שיעמדו בפני מוסדות ביקורת עליונים עם כניסת הבינה המלאכותית, כפי שעולה מדיווחים של רוב מוסדות הביקורת באיחוד האירופי:

- מיומנויות טכניות בלתי מספקות;
- ציות לחוק, לכללי האתיקה, להגנת המידע ולמחויבויות חוזיות;
- המורכבות הרב-תחומית של הנושא;
- היעדר מיומנויות של ניתוח עסקי במוסד;
- מגבלות תקציב.

בתגובה לאתגרים אלה, יש הכרח לקבוע אסטרטגיה מוסדית שתיתן מענה לאתגרים שהוזכרו למעלה ושחייבת ללא ספק לכלול כמה מהאלמנטים הבאים:

1. מודעות ומחויבות של הגופים המנהלים ליישום היוזמות והאמצעים ההכרחיים, הקצאת המימון הנדרש וכוח אדם יעודי לתהליך ההטמעה והפיתוח;
2. מדיניות הנתונים או אסטרטגיית הנתונים בתוך המוסד צריכה להבטיח את זמינותו של מידע עקבי ומהימן כדי שהמוסד יוכל לבצע ניתוח נתונים ולהגן עליהם, וכך למזער נקודות תורפה פוטנציאליות ולספק את הרלוונטיות הדרושה לאמצעי אבטחת הסייבר שיוטמעו;
3. הכשרת כוח אדם כדי לאפשר למבקר להסתגל לסביבה הטכנולוגית והקמת צוותים רב-תחומיים. התוכניות שיפותחו להכשרה ובניית יכולות צריכות להביא בחשבון את השימוש במודלי כשירויות שכבר הוגדרו, כגון זה שנזכר בתרשים 1 אשר תואם לפי הנציבות האירופית למיומנויות הדרושות לעובדים העוסקים בתחום הבינה מלאכותית במגזר הציבורי. על תוכניות אלה לכלול היבטים טכניים של בינה מלאכותית, ובמיוחד ידע על טיפול בנתונים ועיבוד נתונים. ידע זה אמור להיות חלק מהכשרת המבקרים והעובדים העוסקים בטכנולוגיית מידע ותקשורת (ICT) בכל מוסד ביקורת עליון, שכן בדיקת נתונים המשמשים במינהל הציבורי ואיכותם הופכת לחלק חיוני בעבודת הביקורת. ידע מעמיק על שיקולים אתיים, הטיות פוטנציאליות וחקיקה רלוונטית ייחשב גם הוא לרלוונטי;

4. גיוס מומחי אנליטיקת נתונים ומומחים במודלים של בינה מלאכותית, אך גם מומחים בעלי הכשרה משפטית וטכנית, המכשירה אותם לבקר את האלגוריתמים המשמשים במערכות בינה מלאכותית במינהל הציבורי ולפתח פרוטוקולים פנימיים לזיהוי ולצמצום של הטיות במערכות הבינה המלאכותית של מוסד הביקורת עצמו. אלה האחרונים יהיו אמונים על בדיקות סדורות לפני ואחרי הטמעת כלי בינה מלאכותית במטרה להבטיח את ההוגנות והשקיפות של פעולות הפיקוח של המוסד המבקר;
5. שיתוף פעולה בין מוסדות ועבודה שיתופית יחד עם גופי בקרה אחרים וכן קידום חיבור מאובטח בין רשתות פנימיות כדי לשתף מידע שעשוי לסייע לכל מוסד בקרה במשימות של בקרה פנימית וחיצונית על כספי ציבור ופעולות במינהל הציבורי.

4. מסקנות

בשונה ממחקרים רחבים המתמקדים בבינה מלאכותית במינהל הציבורי, עבודה זו מתבוננת בנושא דרך עדשה מוסדית ספציפית המנתחת כיצד מוסדות ביקורת עליונים - המוסדות העצמאיים האמונים על הבקרה על הניהול הכלכלי והתפעול של גופים ציבוריים - מעורבים בבינה מלאכותית הן כמשתמשים והן כמבקרים. האבחון ההשוואתי מציע הבנה מסודרת של המענים הנוכחיים, בדגש על פערים אסטרטגיים והתפתחותן של גישות סדורות (good practices). יש בכך כדי לתרום לפיתוח בסיס ידע משותף, שיכול לתמוך בלמידה הארגונית ובקביעת מדיניות במוסדות ביקורת חיצוניים ברחבי העולם.

הניתוח שבוצע מבליט מספר סוגיות רלוונטיות שמוסדות ביקורת עליונים אינם יכולים להתעלם מהן נוכח העלייה בשימוש בבינה מלאכותית ובנתונים במינהל הציבורי. אף שהתקדמות הבינה המלאכותית בזירה הציבורית מסייעת בשיפור קבלת החלטות, עיצוב שירותים והענקתם, לצד שיפור הניהול הפנימי והיעילות של גופים ציבוריים והשירותים לציבור - היא מציבה אתגרים חדשים עבור מוסדות פיקוח ובקרה. כך, בזמן שמוסדות ביקורת עליונים מספקים פיקוח וביקורת על הציות ורמת הביצוע של שירותים לציבור וגופי ציבור, יהיה עליהם להתמודד עם הסיכונים הנובעים משילוב הבינה המלאכותית - כמו היעדר פרטיות, סודיות ואבטחה של נתוני האזרחים, הצורך בשקיפות בעניין האופן שבו מערכות החלטה אוטונומיות מבוססות אלגוריתמים מחליטות החלטות המשפיעות על חיי אזרחים והצורך להבטיח היעדר אפליה או הטיות באלגוריתמים.

שנית, השימוש במערכות בינה מלאכותית צפוי לעודד אפקטיביות ויעילות בהליכי הביקורת של מוסדות ביקורת עליונים, וזאת כל עוד הם ידעו כיצד לנצל את מגוון הסוגים

של כלי בינה מלאכותית העשויים להועיל להם בתחום פעילותם. כדי להשיג זאת, עליהם להכין נתונים, ומעל לכול, להיות מודעים לחשיבות של הצטיידות בנתונים זמינים ואיכותיים. כפועל יוצא, עליהם להכיר בצורך לבדוק איכות ז, הן במונחי נתוני המגזר הציבורי והן במונחי הנתונים המשמשים את מוסד הבקרה עצמו. יש מוסדות ביקורת עליונים, כמו מוסד ביקורת עליון של ארצות הברית וזה של הודו, שכבר עובדים על כלים המאפשרים להם להשתמש בכמויות גדולות של נתונים כדי לבצע השוואות וניתוחים, ובאמצעותם הם חושפים סיכונים או מצבים שהיה קשה לגלותם בעבר. כלים אלה יכולים למנוע "נקודות של עיוורון" בביקורת ולשפר את היכולת למנוע ולפתור את הסיכונים העיקריים הקשורים לתחומי ממשל שונים. מוסד הביקורת העליון של נפאל משתמש ברשתות עצביות כדי לחלץ אוטומטית מידע קבוע מסוים מתוך מסמכים, ואילו במרבית המקרים שנותרו לעיל מוסדות ביקורת עליונים כבר משתמשים ברובוטים ובמערכות אוטומציה אחרות לבדיקה ובקרה המבוססות על למידת מכונה.

שלישית, הצורך לבקר את מערכות הבינה המלאכותית המשמשות בניהול השירותים לציבור הולך והופך למציאות שכל מוסדות הביקורת העליונים יאלצו להתמודד עימה כדי לשמר אמון ואחריותיות בעידן הדיגיטלי. ננקטו כמה צעדים בנושא, ובכללם פרסומים של מוסדות ביקורת עליונים, כמו דוחות על תמיכה בפיתוח בינה מלאכותית באירופה (בית המשפט האירופי לביקורת), על השימוש בנתונים ובמערכות בינה מלאכותית במשרדי הממשלה (מוסד הביקורת העליון הבריטי), על השימוש במודלים של בינה מלאכותית במינהל הציבורי (מוסד ביקורת עליון של ארצות הברית) ועל כלים מבוססי אלגוריתם בשימוש ממשלתי (מוסד הביקורת העליון של הולנד). אף שההתקדמות בתחום זה איטית בעת הזו, דומה כי תפקידן של הביקורות המבוצעות בידי מוסדות ביקורת עליונים רלוונטי מאוד. זאת, לא רק כדי להבטיח ציות, אפקטיביות ויעילות ביישומן של מערכות בינה מלאכותית, אלא גם כדי לתרום - באמצעות פעילות הביקורת - לפיתוחן של תקנות אסדרה ומדיניות ציבורית שתכליתן לקדם אחריותיות, שקיפות והוגנות של אלגוריתמים.

לסיום, מגבולותיו של מחקר זה נובעות ממספרם המצומצם של הניתוחים הקודמים שבוצעו בנושא מערכות הבינה המלאכותית ופרקטיקות הביקורת של ניהול כלכלי ופיננסי של כספי ציבור ושירותים לציבור. חשוב לשוב ולהדגיש כי עבודה זו מבוססת על ניתוח תפיסתי ואיכותני של התיעוד הקיים. נוכח אופיין המשתנה הן של מערכות בינה מלאכותית והן של ביקורת ציבורית, כמה מההנחות והפרספקטיבות שהוצגו כאן עשויות להצריך תיקוף נוסף. דרוש מחקר נוסף כדי לבדוק אבחנות אלה בשטח ולסגל אותן לפיתוחים חדשים בתחום האסדרתי והטכנולוגי. היעדר נתונים אמפיריים ראשוניים והצורך להסתמך על

tecnologias-y-contabilidad-documentos-aeca/nt18-ps16/ (accessed on 7 February 2025).

3. Wang, P. On Defining Artificial Intelligence. *J. Artif. Gen. Intell.* 2019, 10, 1-37. [CrossRef]

4. McCarthy, J. Why Is Artificial Intelligence? 2004. Available online: https://borghese.di.unimi.it/Teaching/AdvancedIntelligentSystems/Old/IntelligentSystems_2008_2009/Old/IntelligentSystems_2005_2006/Documents/Symbolic/04_McCarthy_whatissai.pdf (accessed on 2 February 2025).

5. Tzafestas, S.G. *Roboethics. A Navigating Overview*; Springer: Cham, Switzerland, 2016.

6. Rich, E. *Artificial Intelligence*; McGraw-Hill: New York, NY, USA, 1983; 411p.

7. Patterson, D.W. *Introduction to Artificial Intelligence and Expert Systems*; Prentice Hall of India: Hoboken, NJ, USA, 1990.

8. Organisation for Economic Cooperation and Development-OECD Artificial Intelligence in Society. 2019. Available online: www.oecd.org/going-digital/artificialintelligencein-society-eedfee77-en.htm (accessed on 7 February 2025).

9. Organisation for Economic Cooperation and Development-OECD Recommendation of the Council on Artificial Intelligence. 2019. Available online: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449> (accessed on 7 February 2025).

10. Fjelland, R. Why general artificial intelligence will not be realized. *Humanit. Soc. Sci. Commun.* 2020, 7, 10. [CrossRef]

11. Berryhill, J.; Heang, K.K.; Clogher, R.; McBride, K. Hello, World: Artificial Intelligence and Its Use in the Public Sector. OECD Working Papers on Public Governance No. 36. 2019. Available online: <https://www.ospi.es/export/sites/ospi/documents/documentos/Tecnologias-habilitantes/IA-Public-Sector.pdf> (accessed on 7 February 2025).

12. Frank, M.R.; Wang, D.; Cebrian, M.; Rahwan, I. The evolution of citation graphs in artificial intelligence research. *Nat. Mach. Intell.* 2019, 1, 79-85. [CrossRef]

מסמכים הזמינים לציבור היו גם הם חיסרון בנייתו של תחום חקר חדש זה. יש לפרש את הממצאים והתובנות מעבודה זו לאור האופי התפיסתי של המחקר. יש להכיר בכך שטכנולוגיית בינה מלאכותית ופרקטיקות של ביקורת המדינה משתנות ומתפתחות לאורך זמן, כך שההנחות והפרספקטיבות שנידונו במאמר זה עשויות להצריך תיקוף אמפירי נוסף או התאמות.

עם זאת, על רקע דברים אלה אנו מעלים הצעות לעבודות מחקר עתידיות, שימשיכו לקדם את השימוש במתודולוגיות אמפיריות מסוימות כמו עבודת שטח מבוססת מקרים, סקרים רחבי היקף וראיונות עם בעלי עניין במוסדות ביקורת עליונים. כל אלה צריכים להתבצע בהקשר של השימוש במערכות בינה מלאכותית ושל ביצוע ביקורות אלגוריתמיות במוסדות ביקורת עליונים. תקוותנו היא כי תוצאות המחקר העתידי בתחום זה יובילו לגיבוש אסטרטגיית בינה מלאכותית, לא רק בקרב מוסדות ביקורת עליונים, אלא בכל מוסד, או אפילו חברה, האמונים על בקרה ופיקוח על הניהול הציבורי. כמו כן, יש לצפות כי המחקר העתידי יוכל לזהות קשיים ומכשולים שבהם עשויים מוסדות ביקורת עליונים להיתקל בשימוש בבינה מלאכותית ולהציע פתרונות טכניים וארגוניים שעשויים לסייע להתגבר עליהם. תחום ניתוח רלוונטי נוסף יכול להתמקד בשימוש פוטנציאלי במערכות בינה מלאכותית בתחום השיפוטי, באותם מוסדות ביקורת עליונים המיישמים פונקציה נוספת זו.

מימון: מחקר זה מומן על ידי משרד המדע והחדשנות הספרדי, מספר מענק/MCIN/128713OB-I00, FEDER (הקרן האירופית לפיתוח אזורי) - "unaManera de Hacer Europa".

תודות: מריה גודז מודה על תמיכה בפריקט המדעי שמספרו PID2023-146575NB-I00, במימון MICIU/AEI/10.13039/501100011033 וכן "ESF".

מקורות

1. Mo Ahn, J. Artificial Intelligence in Public Administration: New Opportunities and Threats. *Korean J. Public Adm.* 2021, 30, 1-33.

2. Asociación Española de Contabilidad y Administración de Empresas (AECA). La transformación digital del sector público en la era del gobierno. In *Documento de las Comisiones Nuevas Tecnologías y Contabilidad (n° 18) y Contabilidad y Administración del Sector Público (n°16)*. 2022. Available online: <https://aeca.es/publicaciones2/documentos/nuevas>

- for Productive, Inclusive and Trusted Governance. OECD Working Papers on Public Governance No. 33. 2019. Available online: https://www.oecd.org/en/publications/a-data-driven-public-sector_09ab162c-en.html (accessed on 7 February 2025).
22. CAF-Banco de Desarrollo de América Latina. Conceptos Fundamentales y Uso Responsable de la Inteligencia Artificial en el Sector Público. Informe 2. Corporación Andina de Fomento. 2022. Available online: <https://scioteca.caf.com/handle/123456789/1921> (accessed on 7 February 2025).
23. Soylyu, A.; Corcho, O.; Elvsaeter, B.; Badenes-Olmedo, C.; Yedro-Martinez, F.; Kovacic, M.; Roman, D. Data Quality Barriers for Transparency in Public Procurement. *Information* **2022**, *13*, 99. [CrossRef]
24. Alhazbi, S. Behavior-Based Machine Learning Approaches to Identify State-Sponsored Trolls on Twitter. *IEEE Access* **2020**, *8*, 195132-195141. [CrossRef]
25. Henrique, B.M.; Sobreiro, V.A.; Kimura, H. Contracting in Brazilian public administration: A machine learning approach. *Expert Syst.* **2020**, *37*, e12550. [CrossRef]
26. Newman, J.; Mintrom, M.; O'Neill, D. Digital technologies, artificial intelligence, and bureaucratic transformation. *Futures* **2022**, *136*, 102886. [CrossRef]
27. Leocadio, D.; Malheiro, L.; Reis, J. Exploration of Audit Technologies in Public Security Agencies: Empirical Research from Portugal. *J. Risk Financ. Manag.* **2025**, *18*, 51. [CrossRef]
28. Cui, I.; Ho, D.E.; Martin, O.; O'Connell, A.J. Governing by Assignment. *SSRN Electron. J.* **2024**, *173*, 157. [CrossRef]
29. Criado, J.I. Inteligencia Artificial (y Administracion Publica). *Econ. Rev. Cult. Leg.* **2021**, *20*, 348-372. [CrossRef]
30. Diakopoulos, N. Accountability in Algorithmic Decision Making. *Commun. ACM* **2016**, *59*, 56-62. [CrossRef]
31. Stone, P.; Brooks, R.; Brynjolfsson, E. 2016 Report, One-Hundred-Year Study on Artificial Intelligence. AI100. 2016. Available online: <https://ai100.stanford.edu/2016-report> (accessed on 7 February 2025).
32. DeSouza, K. Delivering Artificial Intelligence in
13. Reyes Alva, W.A.; Recuenco Cabrera, A.D. Artificial intelligence: Road to a new schematic of the world. *Sciend* **2020**, *23*, 299-308.
14. Bostrom, N. *Superintelligence*; Oxford University Press: Oxford, UK, 2014.
15. Ghosh, M.; Thirugnanam, A. Introduction to Artificial Intelligence. In *Artificial Intelligence for Information Management: A Healthcare Perspective. Studies in Big Data*; Srinivasa, K.G., Siddesh, G.M., Sekhar, S.R.M., Eds.; Springer: Singapore, 2021; Volume 88, pp. 23-44.
16. Xu, C.E.; Xu, L.S.; Lu, Y.Y.; Xu, H.; Zhu, Z.L. E-government recommendation algorithm based on probabilistic semantic cluster analysis in combination of improved collaborative filtering in big-data environment of government affairs. *Pers. Ubiquitous Comput.* **2019**, *23*, 475-485. [CrossRef]
17. Guillem, F. Funciones y Caracterfsticas de la Inteligencia Artificial. *Seguritecnia: Revista decana independiente de seguridad. Seguritecnia* **2022**, *493*, 174-181.
18. Ubaldi, B.; Le Febre, E.M.; Petrucci, E.; Marchionni, P.; Biancalana, C.; Hiltunen, N.; Intravaia, D.M.; Yang, C. State of the Art in the Use of Emerging Technologies in the Public Sector. OECD Working Papers on Public Governance No. 31. 2019. Available online: <https://www.sipotra.it/wp-content/uploads/2019/09/State-of-the-art-in-the-use-of-emerging-technologies-in-the-public-sector.pdf> (accessed on 7 February 2025).
19. Samoil, S.; Lopez, C.M.; Gomez Gutierrez, E.; De Prato, G.; Martinez-Plumed, F.; Delipetrev, B. Defining Artificial Intelligence. Towards An Operational Definition and Taxonomy of Artificial Intelligence (JRC118163) [EUR-Scientific and Technical Research Reports]. Publications Office of the European Union. 2020. Available online: <https://publications.jrc.ec.europa.eu/repository/handle/111111111/59452> (accessed on 7 February 2025).
20. Valle-Cruz, D.; Criado, I.; Sandoval-Almazdn, R.; Ruvalcaba-Gomez, E.A. Assessing the public policy cycle framework in the age of artificial intelligence. From agenda-setting to policy evaluation. *Gov. Inf. Q.* **2020**, *37*, 101509. [CrossRef]
21. Van Ooijen, C.; Ubaldi, B.; Welby, B. A Data-Driven Public Sector: Enabling the Strategic Use of Data

42. Prasad Dotel, R. Artificial Intelligence: Preparing for the Future of Audit. *Int. J. Gov. Audit.* 2020, 47, 32-35. Available online: <https://intosaijournal.org/journal-entry/artificial-intelligence-preparing-for-the-future-of-audit/> (accessed on 8 February 2025).
 43. Garde Roca, J.A. ¿Pueden los algoritmos ser evaluados con rigor? *Encuentros Multidiscip.* 2023, 25, 25. Available online: <http://www.encuentros-multidisciplinarios.org/revista-73/juan-antonio-garde.pdf> (accessed on 8 February 2025).
 44. Koshiyama, A.; Kazim, E.; Treleaven, P.; Rai, P.; Szpruch, L.; Pavey, G.; Ahamat, G.; Leutner, F.; Goebel, R.; Knight, A.; et al. Towards algorithm auditing: Managing legal, ethical and technological risks of AI, ML and associated algorithms. *R. Soc. Open Sci. J.* 2024, 11. Available online: <https://royalsocietypublishing.org/doi/10.1098/rsos.230859> (accessed on 8 February 2025). [CrossRef]
 45. European Court of Auditors. *Artificial Intelligence Initial Strategy and Deployment Roadmap 2024-2025*; European Court of Auditors: Luxembourg, 2024.
 46. National Audit Office. Challenges in Using Data Across Government. UK National Audit Office. 2019. Available online: <https://www.nao.org.uk/insights/challenges-in-using-data-across-government/> (accessed on 8 February 2025).
 47. National Audit Office. Use of Artificial Intelligence in Government. UK National Audit Office. 2024. Available online: <https://www.nao.org.uk/wp-content/uploads/2024/03/use-of-artificial-intelligence-in-government-summary.pdf> (accessed on 8 February 2025).
 48. European Court of Auditors. EU Artificial Intelligence Ambition: Stronger Governance and Increased, More Focused Investment Essential going Forward, Special Report 8/2024. 2024. Available online: <https://www.eca.europa.eu/en/publications/SR-2024-08> (accessed on 8 February 2025).
- הצהרת אי-אחריות/הערת המו"ל: ההצהרות, הדעות והנתונים הנכללים בכל הפרסומים שייכים בלעדית לכל אחד מהמחברים והתורמים למאמר ואינם שייכים ל-MDPI או לעורכים. MDPI ו/או העורכים מסירים אחריות לכל פגיעה באנשים או ברכוש הנגרמת מרעיונות, שיטות, הוראות או מוצרים הנזכרים בתוכן.
- Government: Challenges and Opportunities. IBM Center for Business of Government. 2018. Available online: <http://www.businessofgovernment.org/sites/default/files/Delivering%20Artificial%20Intelligence%20in%20Government.pdf> (accessed on 7 February 2025).
 33. Brookfield Institute. Introduction to AI for Policymakers: Understanding the Shift. 2018. Available online: <https://brookfieldinstitute.ca/intro-to-ai-for-policymakers> (accessed on 7 February 2025).
 34. Kilian, R.; Jack, L.; Ebel, D. European AI Standards-Technical Standardization and Implementation Challenges Under the EU AI Act (26 February 2025). Available online: <https://ssrn.com/abstract=5155591> (accessed on 7 March 2025).
 35. Noble, S.U. *Algorithms of Oppression: How Search Engines Reinforce Racism*; New York University Press: New York, NY, USA, 2018.
 36. European Commission; Joint Research Centre; Medaglia, R.; Mikalef, P.; Tangi, L. *Competences and Governance Practices for Artificial Intelligence in the Public Sector*; JRC138702; Publications Office of the European Union: Luxembourg, 2024; Available online: <https://op.europa.eu/en/publication-detail/-/publication/949913fa-aae4-11ef-acb1-01aa75ed71a1/language-en> (accessed on 7 February 2025).
 37. Filgueiras, F. Inteligencia artificial en la administración pública: Ambigüedad y elección de sistemas de IA y desafíos de gobernanza digital. *Rev. CLAD Reforma Democr* 2021, 79, 5-38. [CrossRef]
 38. Morgan, M.S. Exemplification and the use-values of cases and case studies. *Stud. Hist. Philos. Sci. Part A* 2019, 78, 5-13. [CrossRef]
 39. INTOSAI Development Initiative. *Research Paper on Innovative Audit Technology*; INTOSAI Working Group on Big Data: Beijing, China, 2022.
 40. Amsler, L.B.; Martinez, J.K.; Smith, S.E. *Dispute System Design: Preventing, Managing, and Resolving Conflict*; Stanford University Press: Stanford, CA, USA, 2020.
 41. Rivera, T. Application of machine learning in SAls. *Int. J. Gov. Audit.* 2023, 50, 13-17. Available online: <https://intosaijournal.org/es/journal-entry/machine-learning-application-for-sais/> (accessed on 8 February 2025).