



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 המגזר 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).

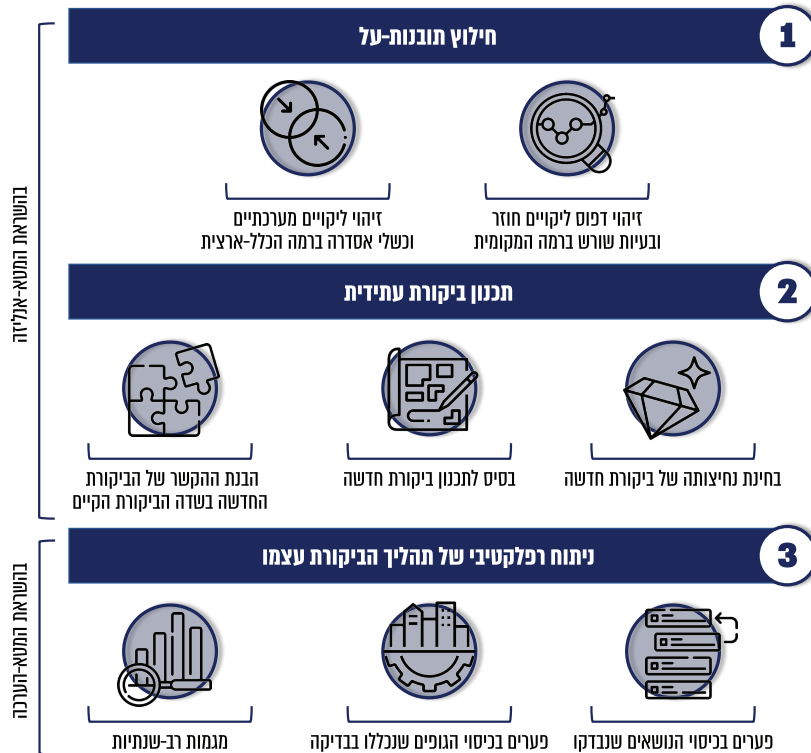
7 שלב מוקדם יותר בהתפתחות המטא-אנליזה הוא שיטת הסקירה הנרטיבית (Narrative Review) שכללה סיכום של ממצאי מחקרים והסקת מסקנה אחת מתכללת מהם.

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 הסתייעות 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 בביקורות 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 בביקורות 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, ושבכולם 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).

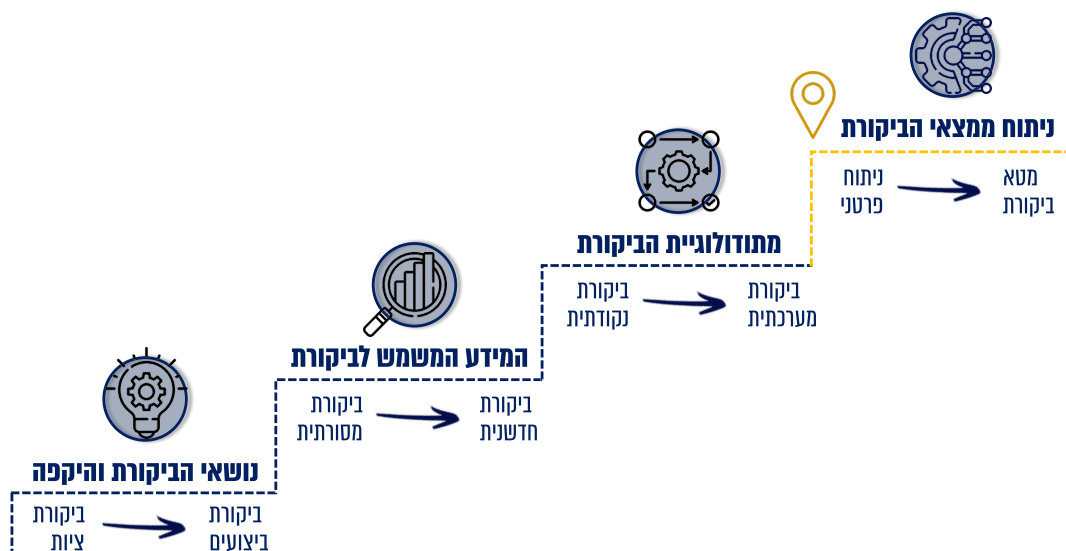
19 ראו ניסים מזרחי "ביקורת המדינה והביקורת הפנימית בישראל - היבטים השוואתיים" 61 **עיונים בביקורת המדינה** 135 (2008); אילת רייכמן-עמית "ללמוד לבחור נכון: מערכת לניהול ובחירה של רשויות מקומיות לביקורת (מנ"ר)" 65 **מראות: עיונים בביקורת המדינה ובבירור תלונות הציבור** 159 (2025).

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 פרידברג עורך 1999).

22 נחום בן-אליא **הדור הרביעי - שלטון מקומי חדש לישראל** (2004); ישי בלנק "מקומו של המ'מקומי: משפט השלטון המקומי, ביזור ואי-שוויון מרחבי בישראל" **משפטים** לד (2) 197 (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 ששימושים מסוימות ניתנות לביצוע ביתר יעילות ולכך שמתפנה 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 לשימושים 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 המחיב. 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).

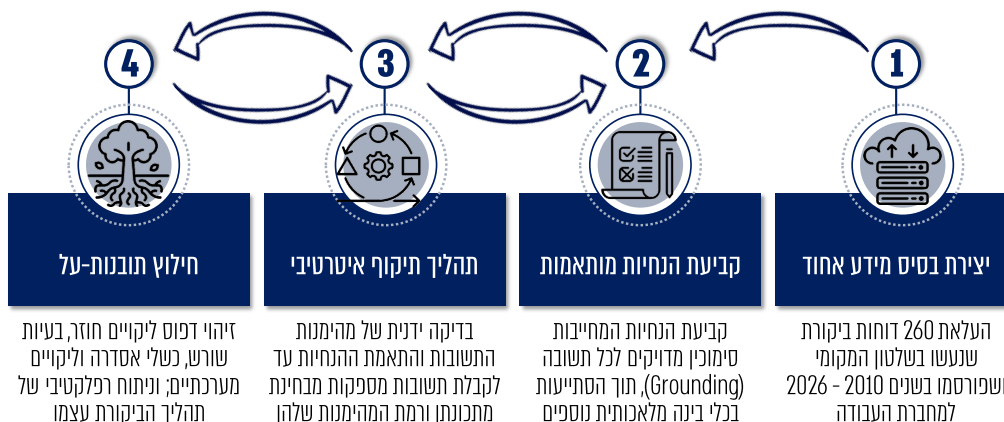
29 המידע הועלה באמצעות היפר-קישור לדוח שפורסם בפורמט PDF. כל דוח הועלה כמקור נפרד. בין המקורות נכללו דוחות רגילים, דוחות מיוחדים (לרבות דוחות ביקורת שפורסמו אחרי פרוץ מלחמת חרבות ברזל) וכן ביקורות מעקב. לא נכללו דוחות ביקורת שכללו אזכור של גופי השלטון המקומי אך לא נכתבו על ידי חטיבת הביקורת על השלטון 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 בפיסוי 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](https://arxiv.org/abs/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 לנותני 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 מציעים 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 קטינים 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 השאלתה 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 מחדשים 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 השולבות זו בזו: (א) סיכונים בשלב שלפית 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 **פלונית 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 המדרג 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 ישיר 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 מסויים. 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 וחיסכון").

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

