

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 להבחין 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 (AI אמין/AI 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 ביקורת? 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>

AI אינו ראייה מעצם הפקתו, וממילא אינו יכול לשמש ממצא מעצם הופעתו. לכל היותר, לאחר עיבוד ביקורתי, אימות, הצלבה ומיקום ביחס לנומרה, אפשר לבחון אם וכיצד תוצר ה-AI עשוי להשתלב במערך ראיות הביקורת. לכן המעבר התקף איננו 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) ¹⁷ אינה תקלה שולית בלבד, אלא דפוס כשל מובנה המחייב מנגנוני אימות, ריסון והימנעות מהסקה ישירה¹⁸. 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 "מנוסח 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 עצמאי¹⁹.

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

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 הנומרה, 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 הנומרה לבין המצב בפועל כפי שהוא עולה מראיות הביקורת¹⁵. מכאן שתוצר

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

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].

39 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 תקף 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,

40 Mökander, *Auditing of AI*, 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 לנזר 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; ומספיקותן 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, לעיל ה"ש 38; 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 אימות, 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 וכתובתו, 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, לארגונו 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 ההנמקה 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 וכתובתו 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.⁵⁴

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 לבחינה 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 החיוביים 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 שנגזרה ממערכת AI בידי גורם אנושי או מנגנון בקרה. אירועים אלה הם חלק מהפיקוח והבקרה על כלי AI, והידע לגביהם הוא חלק מדרישות 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 ותיעוד 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 נשארת 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 בסיסן 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 די; 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 ממוקם 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.