

בינה מלאכותית יוצרת: הזדמנויות, סיכונים ורגולציה

מסמך זה נכתב לקראת דיון ועדת המדע והטכנולוגיה של הכנסת בנושא "בינה מלאכותית: הזדמנויות, סיכונים ושאלת הרגולציה". במסמך מוצג בקצרה נושא הבינה המלאכותית היוצרת שמהווה זרז לשיח הציבורי הער בעת האחרונה; סוגיות מדיניות העולות מהתפתחותה; מידע בדבר החלטות ממשלה, ועדות ותוכנית לאומית בנושא הבינה המלאכותית בישראל, והתייחסות תמציתית למגמות ברגולציה באירופה וארצות הברית.

1. רקע

מסמך של הנציבות האירופית [מגדיר בינה מלאכותית](#) כמערכות המפגינות התנהגות תבונית באמצעות ניתוח סביבתן ונקיטת פעולות, במידה מסוימת של אוטונומיות, כדי להשיג את מטרותן.¹ מערכות בינה מלאכותית יכולות להתבסס על תוכנה בלבד ולפעול במרחב המקוון, למשל בעוזרים קוליים, מנועי חיפוש, מערכות זיהוי פנים, זיהוי דיבור ועוד; או להיות מוטמעות בהתקנים פיזיים כמו: כלי רכב אוטונומיים, רחפנים, רובוטים מתקדמים ועוד.

למרות שיישומי בינה מלאכותית אינם דבר חדש, השילוב בין "קפיצה טכנולוגית" ביכולות העיבוד והניתוח של מערכות כאלה בשנים האחרונות, ותפוצתם של יישומי **בינה מלאכותית יוצרת** (Generative) עוררו לאחרונה דיון ציבורי ער על ההזדמנויות והסיכונים במערכות בינה מלאכותית, [ולקריאה](#) של מובילי טכנולוגיה ומדענים רבים להקפיא למשך חצי שנה פיתוחים חדשים בתחום [ולהצהרה](#) בדבר הצורך לפעול כדי למזער סיכונים חמורים שיכולים להיגרם מטכנולוגיות אלה.

בינה מלאכותית יוצרת היא שם כולל ליישומים המבוססים על מודלי שפה המאפשרים ליצור באמצעות שאילתה קצרה, תוכן חדש - טקסטים, תמונות, סרטונים, מוזיקה, קוד מחשב ועוד, המופק בהתבסס על מידע אימון (training data) שהוזן למערכות. ככל שמידע האימון מפורט, מדויק ומקיף יותר, היכולות של המערכות הללו משתפרות.

מערכות בינה מלאכותית החלו כבר לפני שנים רבות יחסית לבצע בהצטיינות משימות "כמו אנושיות" כגון משחקי שח-מט, משחק הגו הסיני, טריוויה וכדומה.² עם זאת נראה כי פיתוחם של מודלי שפה מתקדמים והשירותים שהם מאפשרים, יוצר שינוי: באינטראקציה אדם - מחשב, בהיקף המשתמשים, ובתפיסה הציבורית.

יישומי בינה מלאכותית יוצרת אינם בגדר "[בינה מלאכותית כללית](#)" (AGI – Artificial General Intelligence), אך כיוון שהם מתמחים במשימות שונות ולא רק במשימה אחת ספציפית וצרה (Narrow AI), הם מעוררים ספק ביחס

¹ להגדרות נוספות ראו גם: רועי גולדשמידט, "[מידע בנושא בינה מלאכותית](#)", מרכז המחקר והמידע של הכנסת, 3 ביוני 2018.

² שם.

להנחות קודמות כי בינה מלאכותית כללית לא צפויה להתממש בעשורים הקרובים וחשש באשר להשלכות המשך הפיתוח והתפוצה הלא מבוקרת שלהם. בעדותו בקונגרס במאי 2023 קרא המנכ"ל של אחת מחברות הבינה המלאכותית לקדם רגולציה שתמרץ בטיחות ונגישות של יישומי בינה מלאכותית, כולל רישוי או רישום של מודלים מעל עוצמה מסוימת ופיתוח מנגנוני פיקוח בינלאומיים על התחום.³

בינה מלאכותית יוצרת היא כבר כיום כלי שימושי ללימוד, הכשרה ויצירה והיצע היישומים בתחום ועוצמתם צפוי להמשיך ולגדול. שימוש מושכל בכלי בינה מלאכותית יוצרת – כזה שאיננו מסתמך עליה באופן בלעדי ובוחן אותה באופן ביקורתי – יכול לייעל ולקצר תהליכי עבודה, למידה ויצירה ולהקנות אפשרויות חדשות למשתמשים. עם זאת, שימוש לא מיומן, לא מבוקר או שימוש לרעה עלולים להוביל לתוצאות שליליות.

טכנולוגיות בתחומי הבינה המלאכותית יוצרות הזדמנויות חדשות ושווקים חדשים בתחומים כגון: בריאות, חינוך אנרגיה, סביבה ועוד. כך לדוגמה, מערכות בינה מלאכותית הסורקות תיקים רפואיים ומידע רפואי, יכולות לזהות סיכונים או סיבוכים צפויים בהליך רפואי ולהציע טיפול; טכנולוגיות ניתוח תמונה העושות שימוש במידע מרשתות חברתיות יכולות לזהות בעלי חיים ספציפיים ואת דפוסי התנועה שלהם; טכנולוגיות ניטור תחבורה מאפשרות לנהל תנועה בכבישים בצורה יעילה וכך לקצר זמני המתנה ולצמצם פליטת מזהמים; מיזמים שונים מנסים להשתמש בנתוני עתק ובניתוח של הידע בהם כדי לפתור סוגיות חברתיות כמו פשיעה, אבטלה או נשירה ממערכת החינוך. מאידך, חדירתן של טכנולוגיות בינה מלאכותית מעוררות גם חששות רבים.

2. בינה מלאכותית יוצרת: סוגיות מדיניות

☒ **השפעות והשלכות על תעסוקה:** בניגוד להנחה המקובלת כי מגמות האוטומציה ישפיעו בעיקר על עבודות כפיים או עבודות "צווארון כחול", המגמות האחרונות בתחום הבינה המלאכותית מצביעות על אפשרויות השפעה גם על משרות בתחומים אחרים בהם: תסריטאים, גרפיקאים, מתכנתים, מתרגמים ועוד. יתרה מכך, לפי [טיוטת מחקר](#) ממרץ 2023 שמנסה לאמוד את ההשלכות של טכנולוגיות בינה מלאכותית יוצרת ומודלי שפה גדולים (LLM's) על שוק התעסוקה בארצות הברית, דווקא מקצועות בשכר גבוה חשופים יותר ממקצועות בשכר נמוך להשלכות חדירתן של טכנולוגיות אלה. לפי אומדן במחקר האמור, בכ-19% מן המשרות צפויה השפעה על כ-50% מן המשימות של העובדים בהן, וב-80% מן המשרות צפויה השפעה על כ-10% מן המשימות של העובדים בהן. עם זאת, לפי [סקר מקיף](#) של הפורום הכלכלי העולמי (WEF) ממאי 2023, שכלל גם חברות ישראליות, שיעור המשרות החדשות שיווצרו בטווח הקצר (2023-2027) צפוי להיות גדול משיעור המשרות שייעלמו, כך שהמאזן הכולל המשוער יהיה חיובי ויוביל לגידול של יותר מ-25% במשרות (בקרוב חברות ישראליות, האומדן של המשיבים הוא כ-11% גידול בהיצע המשרות).

☒ **מידע כזב, "דיפ-פייק" והשלכות על אמון הציבור ועל מוסדות דמוקרטיים:** מודלי שפה חדשים וכלי בינה מלאכותית העושים בהם שימוש, מקלים על יצירת דיסאינפורמציה – תוכן מטעה במתכוון, התאמה שלו לנמענים, והפצתו בהיקפים ניכרים ובקלות.⁴ גישות רווחות לצמצום הבעיה כגון בדיקת עובדות, חינוך משתמשים

³ ראו הרחבה בסוף המסמך בפרק העוסק ברגולציה בארצות הברית.

⁴ ראו גם: רועי גולדשמידט, "[הפצת מידע כזב ותקיפות סייבר לשם השפעה על בחירות](#)", מרכז המחקר והמידע של הכנסת, אפריל 2019.

ושיפור האוריינות שלהם לא בהכרח יעילים כנגד דיסאינפורמציה המפותחת בכלי AI. עם זאת, חברות בתחום עוסקות בפיתוח כלי AI לשם זיהוי ותיוג של תוכן שנוצר באמצעות AI.

☒ **הטיה ואפליה אלגוריתמית:** מערכות מבוססות בינה מלאכותית עושות שימוש רב בבסיסי נתונים קיימים ובמידע המוזן אליהן ומשמש כדי לאפשר להן ללמוד. הטיית בבסיסי הנתונים או בהנחיות המוזנות למערכות כאלה, לצד טעויות מערכת אחרות, עשויות לגרום למערכות כאלה להפלות חלק מן המשתמשים. ככל שחדירתן של מערכות כאלה צפויה לגדול וסמכויותיהן ביחס לקבלת החלטות (בפרט ללא מעורבות אדם) תגדל, עלולה להתרחש "[אפליה אלגוריתמית](#)" בהיקפים נרחבים יותר.

☒ **קניין רוחני והשלכות על יצירה:** מסמך שפרסמו לאחרונה [שירותי המחקר של הקונגרס בארצות הברית](#) מעלה שאלות מעניינות ביחס לנושא הקניין הרוחני, בהן: האם ניתן לקבוע זכויות יוצרים עבור תוכן שהופק באמצעות מערכות AI? האם ניתן להגדיר להן "יוצר" או "מחבר" (Authorship)? באם ניתן לרשום זכויות יוצרים, למי מוקנות הזכויות (לחברת התוכנה? למפתח? למשתמש הקצה?) כיוון שכאמור לעיל, יישומי בינה מלאכותית זקוקים לכמויות עצומות של מידע אימון, האם השימוש במידע אימון שיש עליו זכויות יוצרים נחשב להפרה שלהן? או לשימוש הוגן? והאם תוצרי בינה מלאכותית (כגון טקסטים או תמונות) נחשבים למפרים קניין רוחני במקרה של דמיון משמעותי ושל ראייה לכך שהתוכן המוגן שימש לשם אימון המודל.

הזיות של בינה מלאכותית הוא מושג מדעי מקובל לתיאור מקרים בהם מערכות בינה מלאכותית מספקות טקסט, תמונה או תוכן אחר - חסר משמעות, מוטעה או לא רלבנטי בהקשר הנידון. מצב זה פוגע במהימנות של מערכות אלה ובציפיות המשתמשים מן המערכות ומהווה אחד מן האתגרים בהבשלתן של טכנולוגיות אלה לכדי מוצרים אמיתיים. במקרים של שימוש במערכות כאלה בתהליכי החלטה (למשל ביחס למתן תרופה, לביצוע הליך רפואי, או במקרה של כלים אוטונומיים) לטעויות מערכת מסוג זה יכולות להיות השלכות כבדות משקל, בפרט כיוון שלא תמיד קל לזהות אותן. להרחבה [ראו](#).

3. בינה מלאכותית: מדיניות ממשלתית בישראל

[החלטת ממשלה 212 מאוגוסט 2021](#) הטילה על שרת החדשנות, המדע והטכנולוגיה להוביל את מדיניות הממשלה בתחום הבינה המלאכותית בנושאי רגולציה, מדיניות מידע ונתונים, אתיקה, שיתוף פעולה בין-לאומי אזרחי והטמעה במגזר הציבורי האזרחי ועוד. לשם מימוש מטרות אלה, הוחלט להקים צוות בין-משרדי בהובלת מנכ"לית משרד החדשנות, המדע והטכנולוגיה שיעסוק בנושא הבינה המלאכותית ויגיש המלצות בתחום. עוד נקבע בהחלטה, כי יש לאשר את מתווה פורום תל"מ (תשתיות לאומיות למדע) לבינה מלאכותית ומדעי הנתונים בהתאם להמלצות הוועדה שמינה הפורום (ראו להלן). המלצות מתווה תל"מ כוללות: "פיתוח הון אנושי מחקרי בתחומי הליבה של טכנולוגיית הבינה המלאכותית; הקמת מרכז חישוב על-לאומי לשימוש המגזר הציבורי הרחב, התעשייה והאקדמיה; קידום יכולות עיבוד שפה טבעית בעברית ובערבית; יצירת סביבה רגולטורית מאפשרת להתפתחות תחומים אלה, ועוד".

ביולי 2022 [פורסם](#) כי ישראל גיבשה "תוכנית לאומית לבינה מלאכותית" בהובלה של משרד החדשנות, המדע והטכנולוגיה ורשות החדשנות הפועלת במסגרת המשרד. התוכנית הלאומית האמורה, גובשה בעקבות עבודה של [ועדת בינה מלאכותית ומדע הנתונים](#) שהגישה את מסקנותיה ליו"ר פורום תל"ם בדצמבר 2020.⁵

עוד טרם עבודת הוועדה האמורה, בוצעו עבודות גם על ידי מומחים נוספים, בהן דוח בנושא "[אתיקה ורגולציה של בינה מלאכותית](#)" שנכתב כחלק מהמיזם הלאומי למערכות נבונות ובטוחות, ודוח "[המיזם הלאומי למערכות נבונות בטוחות להעצמת הביטחון הלאומי והחוסן המדעי-טכנולוגי: אסטרטגיה לאומית לישראל](#)". צוות המיזם החל לעסוק בנושא בשנת 2018, אך הדוח פורסם רק ב-2020. יצוין, כי **במתווה שהוצע במסגרת המיזם הלאומי האמור, הומלץ לתקצב את המיזם ב-10 מיליארדי ש"ח למשך 5 שנים, ולהקים מינהלת לאומית ייעודית לניהולו במשרד ראש הממשלה. לעומת זאת, בדוח ועדת בינה מלאכותית ומדע הנתונים, גובה ההקצאה הנדרש היה נמוך משמעותית - כ-5.2 מיליארד ש"ח, והוצע לבססו על גופים קיימים.**

יצוין כי למרות שכאמור פורסם כי גובשה "תוכנית לאומית לבינה מלאכותית" על בסיס המלצות פורום תל"ם, לא מצאנו פרסום מלא של מתווה התוכנית ולא ברור מה סך התקציב המיועד לתוכנית. לפי ד"ר זיו קציר, מנהל תוכנית תל"ם לבינה מלאכותית, **ההשקעה הכוללת במתווה עומדת על מיליארד ש"ח בשתי פעימות** (בתקציבים שאושרו עד כה, כולל תקצוב 2023-2024).

לטענת גורמים שנטלו חלק בתכנון המיזם הלאומי למערכות נבונות, היקפה המוגבל של התוכנית הלאומית שהתקבלה פוגע ביכולתה של ישראל למצב עצמה כשחקן מוביל, משפיע ובעל כוח שוק בתחום הבינה המלאכותית – בדומה לעוצמה של ישראל בתחום הסייבר, הנובעת מהזיהוי המוקדם וריכוז המאמץ בתחום הסייבר. **ראש הממשלה, חה"כ בנימין נתניהו התייחס לנושא ב-5 ביוני 2023 וציין כי יכנס בימים הקרובים צוותי חשיבה כדי לדון על המדיניות הלאומית של מדינת ישראל בנושא הבינה המלאכותית, הן בתחום האזרחי והן בתחום הביטחוני.** זאת במטרה להפוך את ישראל למעצמה לאומית בתחום הבינה המלאכותית, כפי שהיא בתחום הסייבר.⁶ גיבוש המדיניות הלאומית שהתקבלה נעשה בין השאר על רקע התיאור המופיע בדוח ועדת בינה מלאכותית ומדע הנתונים לפיו קיים בישראל פער ניכר בין המחקר והפיתוח (להלן: המו"פ) המתקדם בתחומי הבינה המלאכותית לבין חסר באסטרטגיה ממשלתית בתחום שבא לידי ביטוי גם במחסור בתשתיות חישוב הולמות. כותבי הדוח הגיעו למסקנה כי **"קיים צורך קריטי בייזום תוכנית לאומית שתתווה פתרון מערכתי לקידום המו"פ בתחום בינה מלאכותית ומדע הנתונים. יש למדינת ישראל את הפוטנציאל וההזדמנות להיות מדינה מובילה בתחום. קיומה של תוכנית לאומית בתחום בינה מלאכותית ומדע הנתונים, חיוני לחוסנה של מדינת ישראל ויתרום להעמקת המחקר האקדמי, להרחבת הבסיס התעשייתי, ליצירת הזדמנויות כלכליות משמעותיות ולקידום דרמטי של היכולות הביטחוניות הישראליות."**

באוקטובר 2022 פרסמה שרת החדשנות, המדע והטכנולוגיה דאז מסמך מדיניות להערות הציבור תחת הכותרת "[מדיניות רגולציה ואתיקה בתחום הבינה המלאכותית](#)".

⁵ הגרסה הראשונה של הדוח מדצמבר 2020 זמינה [באתר רשות החדשנות](#).

⁶ בנימין נתניהו, ראש ממשלת ישראל, בערוץ [הטוויטר הרשמי](#), כניסה: 6 ביוני 2023.

▪ לפי המסמך "אין מקום בעת הזו לקידום רגולציה באמצעות חקיקה רוחבית שהיא ייחודית לתחום הבינה המלאכותית. יוער כי ייתכן ומדיניות זו תשתנה ככל שתחום הבינה המלאכותית יתפתח ויתבסס ומשמעויותיו תתבהרנה".

▪ בין העקרונות האתיים המפורטים במסמך: בינה מלאכותית תשמש לקידום צמיחה, פיתוח בר-קיימא והמובילות הישראלית בתחום החדשנות; האדם במרכז: כיבוד זכויות יסוד ואינטרסים ציבוריים; שוויון ומניעת אפליה פסולה; שקיפות והסברתיות; אמינות, עמידות, אבטחה ובטיחות, ו-אחריותיות.

▪ בנייר עמדה של המכון הישראלי לדמוקרטיה בנושא, נטען בין היתר כי: (1) הגישה במסמך המדיניות הממשלתי לפיה די באסדרה רכה ואתית בלבד לא עולה בקנה אחד עם ההתפתחויות בעולם ותוביל לפתיחת פער מיותר בין ישראל למדינות המפותחות; (2) המסמך מתעלם מסוגיית שימושי בינה מלאכותית במגזר הציבורי ומתמקד בהסדרת השוק הפרטי; (3) ההתייחסות להסדרת בינה מלאכותית כשלעצמה, בהתעלם מהסדרת טכנולוגיות משיקות, כמו פרטיות וסייבר והכשל בהסדרתן - היא בעייתית.

החלטת ממשלה 173 מפברואר 2023 תחת הכותרת "חיזוק המובילות הטכנולוגית של ישראל" מאשרת את

התוכנית הלאומית לבינה מלאכותית ובכללה: האצת המחקר הבסיסי והיישומי בתחום; יצירת קפיצת מדרגה לתעשייה הישראלית בפיתוח תשתיות לתחום והטמעה של יישומי בינה מלאכותית במגזר הציבורי האזרחי לייעול. לפי ההחלטה: משרד האוצר יקצה תקציב הרשאה להתחייב שלא יעלה על 500 מיליון ש"ח, והגופים החברים בפורום תל"מ יעמידו את תקציב המתווה המיועד להקצאה על-ידם והוא ימומש בשנים 2023 עד 2026. כאמור, לפי ד"ר זיו קציר, מנהל תוכנית תל"מ לבינה מלאכותית, **ההשקעה הכוללת במתווה עומדת על מיליארד ש"ח בשתי פעימות**.⁷

יובהר, כי לא סקרנו במסמך את פעילות הממשלה בפועל בתחומי הבינה המלאכותית, אך בין השאר אושרה תמיכה בכ- 5.5 מיליוני ש"ח **ליצירה והנגשה של מאגרי מידע בעברית וערבית** שאמורים לשמש תשתית למודלי שפה; פורסם **קול קורא להקמת מעבדת מו"פ** בתחומי חישוב על ובינה מלאכותית; פורסם קול קורא **לפיתוח כלי תוכנה להסברתיות, שימור פרטיות ולמידה מבוזרת**, הוקם **צוות בין משרדי** לבחינת השימוש בבינה מלאכותית במגזר הפיננסי; ונחתם מזכר **לשיתוף פעולה בין הכנסת לרשות החדשנות** בתחומי עיבוד שפה טבעית, על בסיס הנגשת מאגרי מידע של הכנסת.

4. בינה מלאכותית: מידע ראשוני על יוזמות רגולציה בעולם

בעוד כאמור לעיל, העמדה הממשלתית הרשמית האחרונה שפורסמה בנושא תמכה ברגולציה רכה ולא ביוזמת חקיקה מקיפה לנושא, באירופה קיימת טיוטת חוק מקיפה בנושא בינה מלאכותית. לפי **פרסום** של הפרלמנט האירופי עיקרי טיוטת החקיקה כוללים הבחנה בין רמות שונות של סיכון שמציבות מערכות בינה מלאכותית. כך למשל, מערכות ברמת סיכון גבוהה מאוד לבני אדם, בהן: מערכות המשתמשות במסרים מניפולטיביים או תת-הכרתיים, מנצלות בני אדם או משמשות ל"דירוג חברתי" (לפי התנהגויות, מצב כלכלי-חברתי ועוד), ייאסרו בשימוש.

⁷ ד"ר זיו קציר, מנהל תוכנית תל"מ לבינה מלאכותית, רשות החדשנות, דוא"ל מענה לפניית מרכז המחקר והמידע של הכנסת, 1 ביוני 2023.

שימושים אסורים נוספים המוצעים בטיטת החקיקה, כוללים גם: זיהוי ביומטרי מרחוק בזמן אמת במקומות ציבוריים; זיהוי ביומטרי מרחוק למפרע, למעט עבור רשויות איפה ובנסיבות של פשעים חמורים ובכפוף לאישור משפטי; מיון ביומטרי המתבסס על מאפיינים רגישים (כגון מגדר, אתניות, אזרחות, דת או עמדה פוליטית); "מערכות שיטור מנבא" (Predictive policing); מערכות "זיהוי רגשות" לאכיפת חוק, לניהול גבולות, במקום העבודה ובמוסדות חינוך ועוד.

על פי הטיטה, מערכות המוגדרות ככאלה שיש בהן סיכון גבוה כוללות כאלה שמזיקות לבריאות, לבריאות או לזכויות היסוד של בני אדם או לסביבה; מערכות בינה מלאכותית שמשפיעות על בוחרים בקמפינים פוליטיים; ו"מערכות המלצה" שמשמשות רשתות חברתיות שלהן 45 מיליוני משתמשים ומעלה.

עוד על פי הטיטה, ספקים של מודלי בסיס (Foundation Models) בתחום הבינה המלאכותית יצטרכו להתחייב להגן על זכויות יסוד, בריאות, ביטחון, הגנת הסביבה, הדמוקרטיה ושלטון החוק; לאמוד ולמתן סיכונים; לעמוד בדרישות שונות ואף לרשום את המודלים במאגר אירופי. מודלים בסיסיים יוצרים כמו GPT יידרשו בדרישות שקיפות נוספות בהן: חובת גילוי שהתוכן נוצר על ידי בינה מלאכותית; עיצוב המודל כך שלא יצרוך תוכן לא חוקי ולא יפרסם סיכומים של מידע או נתונים מוגנים בזכויות יוצרים ששימשו לאימון המודל. עם זאת, כדי לעודד חדשנות, מוצע להחריג מן החוק פעילויות לשם מחקר ורכיבי בינה מלאכותית שמונגשים תחת רישיון בקוד פתוח. החוק אמור לעודד סביבות ניסוי נשלטות ו"קופסאות חול רגולטוריות" שישמשו לבחינה מוקדמת טרם פריסתם.

"בעיית ההתאמה" (alignment) הוא מושג שנועד לתאר את האתגר והקושי בהתאמה בין "התנהגותן" או פעולתן של מערכות בינה מלאכותית לבין נורמות וערכים אנושיים. בעיה זו הופכת מרכזית יותר ככל שמידת האוטונומיות ומהירות הפעולה של המערכות גדלה ומונעת יכולת פיקוח והערכה אנושית של פעולתן. בעיית ההתאמה כוללת הן היבט טכני: כיצד להטמיע בקוד המערכות ערכים ועקרונות כך שהן יפעלו כמצופה מהן? והן היבט נורמטיבי: מה הם הערכים והעקרונות שיש להטמיע במערכות? דוגמה אחת לבעיית ההתאמה, התרחשה בפועל כאשר רובטי שיחה (צ'ט בוט) שהורשו לנהל שיח חופשי עם אנשים באינטרנט החלו לאמץ שיח פוגעני (על בסיס למידה מהמשתמשים); חשש אחר הוא ממצב בו מערכות בינה מלאכותית חזקות, יבחרו בשיטות יצירתיות או חדשניות להשגת המטרה שהוגדרה להן, בדרכים שהן פוגעניות או מזיקות לאנושות.⁸

בשונה מהעמדה הכוללנית והאקטיבית המסתמנת בכיווני הרגולציה של האיחוד האירופי, עד כה, טרם גובשה מסגרת או תפיסה דומה ברגולציה בארצות הברית (בהיבט זה, יתכן כי המצב דומה לגישות השונות להגנת הפרטיות בין אירופה לבין בארה"ב), אף כי גם בארצות הברית מתקיימים דיונים בנושא הצורך ברגולציה על בינה מלאכותית.

ארצות הברית

בפברואר 2019 פרסם הבית הלבן "צו נשיאותי" (Executive Order) תחת הכותרת "[שימור ההובלה האמריקאית](#) [בבינה מלאכותית](#)". הצו קובע חמישה עקרונות: (1) להוביל פריצות דרך טכנולוגיות בבינה מלאכותית בכלל המגזרים: ממשל פדרלי, תעשייה ואקדמיה כדי לקדם גילויים מדעיים, תחרותיות כלכלית וביטחון לאומי; (2) להוביל פיתוח

⁸ Gabriel Lason, "[Artificial Intelligence, Values, and Alignment](#)", *Minds and Machines* (30), 2020, pp.411–437.

סטנדרטים ולהסיר חסמים לבדיקה ויישום של טכנולוגיות בינה מלאכותית; 3) להכשיר את העובדים למיומנות בפיתוח והטמעה של טכנולוגיות בינה מלאכותית כדי להכשיר אותם להווה ועתיד התעסוקה; 4) לטפח את אמון הציבור ובטחונו בטכנולוגיות בינה מלאכותית, ולהגן על חירויות אזרחיות, פרטיות וערכים אמריקאים; 5) לקדם סביבה בינלאומית התומכת במחקר וחדשנות אמריקאית בתחום הבינה המלאכותית.

בהמשך לצו הנשיאותי האמור, פרסם המשרד הנשיאותי לניהול ותקציב (OMB) בנובמבר 2020 מזכר [הנחיות לרגולציה על יישומי בינה מלאכותית](#) לראשי הסוכנויות. בין השאר נכתב במזכר כי על הסוכנויות להימנע מרגולציה או פעולות אחרות שיסכנו את החדשנות והצמיחה של תחום הבינה המלאכותית ולהימנע מלנקוט גישה זהירה שתגדיר למערכות בינה מלאכותית סטנדרט גבוה ובלתי אפשרי שימנע מהחברה לנצל את היתרונות שלה, ושיכולה לפגוע במעמדה של ארצות הברית כמובילה בתחום הבינה המלאכותית. המזכר מתייחס לשורת עקרונות מנחים לעיצוב רגולציה בהם: אמון הציבור במערכות בינה מלאכותית; שיתוף ציבור, יושרה מדעית ואיכות המידע; הערכה וניהול סיכונים; תועלות ועלויות; גמישות; הוגנות ואי-אפליה; גילוי נאות ושקיפות, בטיחות ובטחון ושיתופי פעולה בין סוכנויות. לפי [סקירה](#) של מכון ברוקינגס בנושא, הגישה במזכר משקפת מדיניות של צמצום רגולציה מחד, והסמכתן של הרשויות והסוכנויות הפדרליות לפעול בתחומי אחריותן בכלים רגולטוריים ולא רגולטוריים בהתאם לגישה מבוססת רמת סיכון. למרות שהמזכר דרש מהסוכנויות הממשלתיות הרלבנטיות לקבוע תוכנית לרגולציה של בינה מלאכותית, רק חלק קטן מ-41 הסוכנויות הרלבנטיות עשו זאת עד כה.

באוקטובר 2022 פרסם הבית הלבן תוכנית ל"[הצעת חוק זכויות בבינה מלאכותית](#)" (AI bill of rights), המתבססת על חמישה עקרונות: מערכות בטוחות ואפקטיביות; הגנה מפני אפליה אלגוריתמית; הגנת מידע; התרעה והסבר וחלופות אנושיות. מן הסקירה של מכון ברוקינגס עולה כי התוכנית האמורה מתמקדת בצמצום הסיכונים לכלכלה ולזכויות אדם עקב מערכות בינה מלאכותית – בשונה מהמיקוד של תוכניות ממשלתיות קודמות בהזדמנויות לחברה ולכלכלה. כמו כן, בשלב זה התוכנית היא בגדר המלצה בלבד ויש הבדלים במידת יישומה בסוכנויות פדרליות שונות. יצוין כי בינואר 2023 [פרסם](#) המכון הלאומי לסטנדרטים וטכנולוגיה (NIST) מסמך מסגרת להערכת סיכונים בבינה מלאכותית.

במאי 2023 העיד מנכ"ל חברת הבינה המלאכותית OpenAI סם אלטמן בפני ועדת משנה של הסנאט לפרטיות, טכנולוגיה וחוק. בעדותו ציין אלטמן כי לשיטתו רגולציה על בינה מלאכותית היא חיונית כדי לתמרץ בטיחות של מערכות בינה מלאכותית לצד ההגנה על אפשרותם של אנשים להשתמש בהן ולהפיק מהן תועלת. כדי לוודא את השמירה על בטיחות, בפרט בחברות המפתחות מודלים חזקים של בינה מלאכותית, יש לוודא כי הן עומדות בדרישות בטיחות הולמות, כולל בדיקות פנימיות וחיזונית בטרם פרסום. הוא קרא לממשלת ארצות הברית לשקול שילוב של רישוי או רישום של מודלי בינה מלאכותית מעל רף ביצועים מסוים, לצד תמריצים לגופים שיעמדו בדרישות האמורות. עם זאת, בשל הדינמיות של תחום הבינה המלאכותית, נדרש משטר רגולטורי גמיש מספיק כדי להתאים להתפתחויות הטכנולוגיות ולבחון הליך שמשלב את כלל בעלי העניין כדי לפתח ולעדכן תדיר סטנדרטים של בטיחות, תנאי הערכה, כללי חשיפה, תהליכי תיקוף. בנוסף, לטענתו על מעצבי מדיניות לבחון כיצד להטמיע רגולציית רישוי

ברמה גלובאלית כדי לוודא שיתופי פעולה בינלאומיים בנושא בטיחות בבינה מלאכותית ואף לבחון הקמת ארגון - בין-ממשלתי לפיקוח והגדרת סטנדרטים.⁹

יצוין כי נושא הבינה המלאכותית נדון כיום בקונגרס האמריקני בהקשרים שונים ובוועדות שונות, ובתוך כך מתקיימים דיונים על הצורך ברגולציה על תחום זה. כך למשל, בדיון של [ועדת הסנאט לביטחון לאומי](#) (Homeland Security) בו נדונו היתרונות והסיכונים של השימוש בטכנולוגיות אלו על ידי סוכנויות הממשל והצורך בהכשרה הולמת לעובדיהן.

⁹ Written Testimony of Sam Altman, Chief Executive Officer OpenAI Before the U.S. Senate Committee on the Judiciary Subcommittee on Privacy, Technology, & the Law, "[Oversight of A.I.: Rules for Artificial Intelligence](#)", May 16th, 2023.