

פרק 1

מהי בינה מלאכותית?

המונח **בינה מלאכותית** מיוחס לג'ון מקארתי, מדען מחשב וחוקר בתחום המדעים הקוגניטיביים, שבשנת 1956 ארגן את הכנס האקדמי הראשון בנושא. תחילת הדרך מעט מוקדמת יותר ובין האבות המייסדים מקובל למנות את וניבר בוש (Vannevar Bush), שכבר בשנת 1945 הציע מערכת שתגביר ידע והבנה אנושיים, ואת אלן טיורינג, שבשנת 1950 כתב מאמר על מסוגלות של מכונות לדמות בני אדם ועל יכולתן לבצע פעולות אינטליגנטיות דוגמת משחק שחמט.¹ מי שנחשב גם הוא בין האבות המייסדים הוא מרווין לי מינסקי, שהוכשר כד"ר למתמטיקה והיה מעורב במחקרים, בהמצאות ובפיתוחים רבים בתחום. מינסקי הוא שטבע את ההגדרה הפופולרית לבינה מלאכותית: "לגרום למכונה להתנהג בדרך שהייתה נחשבת לאינטליגנטית לו אדם התנהג כך".² בתחילת הדרך של חקר הבינה המלאכותית שלטה הפרדיגמה ה"סימבולית", ששאפה לשכפל חשיבה אנושית ברמה גבוהה. עם השנים היא הוחלפה על ידי פרדיגמה "חיבורית", המבקשת לחקות את הבסיס הביולוגי של הקוגניציה האנושית באמצעות רשתות נוירונים מלאכותיות. עם זאת, במאה ה-20 הפרדיגמות הללו לא מימשו ציפיות שהיו תלויות בהן מעבר להדגמות תאורטיות או מעבדתיות, דבר שהוביל למה שמכונה "חורף הבינה המלאכותית", שבמסגרתו צומצמו המחקר וההשקעות בזירה זו לתקופות ארוכות.³

בעשור האחרון, עקב התקדמות בחקר מדעי המחשב וכן בפיתוח חומרה ותוכנה בתחומים של מחשוב ותקשורת, יחד עם פריצת דרך במחשוב ענן ונתוני עתק, מתאפשרת התקדמות משמעותית בבינה המלאכותית ובתוכה תת-תחומים דוגמת למידת מכונה ורשתות נוירונים עמוקות (מושגים אלו ייסקרו בהרחבה בהמשך). מחקרים מסוימים אף טוענים כי ההתקדמות ברשתות נוירונים עמוקות משמעותית עד כדי כך שתת-תחום זה נחשב כמעט לשם נרדף למונח בינה מלאכותית.⁴

מרבית היישומים הנפוצים בבינה המלאכותית כיום שייכים לתת-תחום שנקרא למידת מכונה, הכוללת אלגוריתמים סטטיסטיים שמבקשים לחקות משימות אנושיות קוגניטיביות על ידי גזירת חוקים בנוגע אליהן, באמצעות ניתוח כמויות גדולות של

נתונים. האלגוריתם למעשה "מתאמן" על מידע קיים ובמהלך האימון יוצר מעין מודל סטטיסטי משלו, על מנת לבצע את אותה משימה בעתיד על נתונים חדשים שבהם לא נתקל קודם לכן.⁵

בינה מלאכותית שייכת לתחום רחב יותר של מדעי הנתונים, ואכן נחוצים לה נתונים רבים על מנת לפעול בצורה יעילה. מדובר כיום על הצורך בנתוני עתק (Big Data), אשר מספקים זרם סדיר של נתונים ההכרחיים לגזירת תובנות משמעותיות בעזרת אלגוריתמים לומדים. אולם הבינה המלאכותית אינה תלויה רק בנתוני עתק, אלא היא אחד האמצעים היעילים היחידים להפקת ערך וידע מכמות כזו של נתונים, שלצורך ניתוחם נדרשים אלגוריתמים חזקים במיוחד.⁶

חלק ניכר מעבודת האבות המייסדים של הבינה המלאכותית הוא הבסיס התאורטי לאלגוריתמים של למידת מכונה, הנמצאים בשימוש במערכות רבות בנות זמננו ומאפשרים פעולות ובהן זיהוי תמונה ונהיגה אוטונומית.⁷ מערכות אלו שייכות למה שמקובל לכנות בינה מלאכותית צרה או בינה מלאכותית חלשה, וזאת למרות שלעיתים מדובר ביישומים מתקדמים. הכוונה במושג זה היא לאלגוריתמים שנועדו להתמודד עם מקבץ של בעיות ספציפיות וביניהן למשל משחק, זיהוי תמונה או ניווט.⁸ מושג זה שונה מבינה מלאכותית כללית, המתייחסת למערכת המסוגלת להשתמש באינטליגנציה ברמה אנושית במגוון רחב של משימות.⁹ בינה מלאכותית כללית עדיין אינה קיימת, נכון לכתבת שורות אלו, והדעות חלוקות בנוגע ליכולת ליצור כזו לפחות בשני העשורים הקרובים. הבינה המלאכותית שמתפתחת כיום היא בעיקר יישומי למידה עמוקה. טכנולוגיה זו אומנם שייכת לקטגוריה של בינה מלאכותית צרה, אולם היא מאפשרת צורה מדויקת יותר של למידה ממוחשבת ומנגישה שימוש מסחרי נרחב יותר ביישומי בינה מלאכותית.¹⁰

רקע היסטורי: שלושת הגלים הראשונים בבינה המלאכותית

מקובל לחלק את ההתפתחות בבינה המלאכותית לשלושה גלים, בהתבסס על ניתוח התפתחות היכולות של התחום. הסוכנות הצבאית לפרויקטים מחקרניים מתקדמים (להלן דארפ"א)¹¹ במחלקת ההגנה של ארצות הברית (DoD) היא אחד הגורמים המובילים בעולם בפיתוח בינה מלאכותית לצרכים ביטחוניים. הסוכנות מגדירה בינה מלאכותית כ"יכולת מתוכנתת לעבד מידע",¹² אולם לצד ההגדרה הפשוטה הזו גם שם מחולק הנושא לשלושה גלים, כאשר כל גל מאופיין על 'סולם אינטליגנציה רעיוני' (Notional Intelligence Scale) שבמסגרתו נמדדת כל אחת מארבע היכולות הבאות בדומה לממדים של אינטליגנציה אנושית:

- חישה (Perceiving) – היכולת להבחין בהתרחשויות בעולם;
- למידה (Learning) – היכולת ללמוד דברים ולהסתגל לסיטואציות שונות;
- הפשטה (Abstracting) – היכולת לקחת ידע שהתגלה ברמה מסוימת ולהשליך או ליישם אותו על רמה אחרת;
- הנמקה הגיונית (Reasoning) – היכולת לנמק באופן הגיוני או לקבל החלטות בנוגע לדברים באופן הגיוני.

גל ראשון – תיאור 'ידע בעבודת יד'

הגל הראשון מבוסס על 'ידע בעבודת יד' (Handcrafted knowledge), ובו אספו מומחים ידע קיים בנושא מסוים ואפיינו אותו במסגרת חוקים שיכלו להתאים למחשב, שבתורו יכול היה ללמוד את ההשלכות של החוקים הללו.¹³ לדור זה של בינה מלאכותית שייכות למשל תוכנות לוגיסטיות לתכנון פעולות כגון משלוחים, תוכנות לחישוב מיסים ותוכנות שהצליחו לשחק שחמט נגד בני אדם. תוכנות ויישומים ממוחשבים רבים שבהם נעשה שימוש כיום מבוססים על הגל הראשון של הבינה המלאכותית, והתפיסות העומדות בבסיסו עדיין רלוונטיות לשימושים צרים, כפי שקורה בחלק ניכר מן האפליקציות בטלפונים החכמים או בתוכנות דוגמת Office. הדור הזה של הבינה המלאכותית אינו מצטיין בלמידה, אולם לטענת דארפ"א הוא ממשיך להתפתח ולהשיג הישגים רבים, למשל בתחומי הגנת הסייבר, והוא עדיין מאוד רלוונטי גם בימינו.¹⁴ מבחינת דארפ"א, תוצרי הגל הראשון הם בעלי יכולת חישה בינונית ויכולת להסביר סיבתיות בהיבטים צרים מאוד, אולם הם חסרי יכולת למידה ויש להם יכולת מועטה מאוד להתמודד עם חוסר ודאות.

בינה מלאכותית מהגל הראשון



מאפשר הנמקות הגיוניות עבור בעיות מוגדרות ומתחמות.

ללא יכולות למידה ויכולת דלה להתמודד עם חוסר ודאות

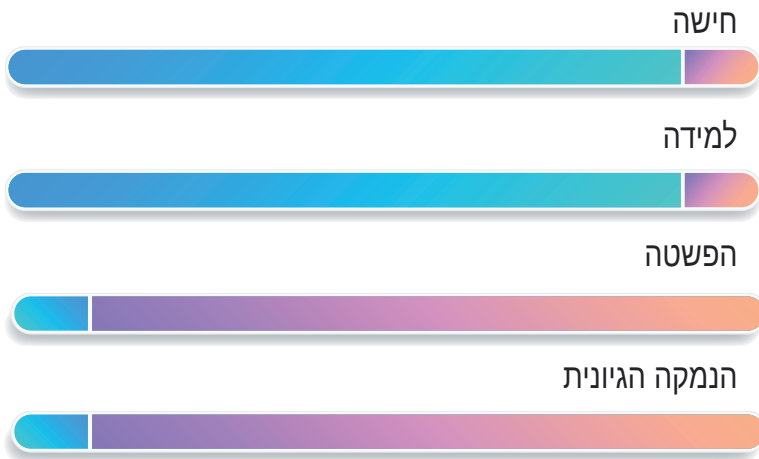
¹⁵ מבוסס על שרטוט מתוך: A DARPA Perspective on Artificial Intelligence

גל שני – 'קטגוריזציה': למידה סטטיסטית

הגל השני הוא 'קטגוריזציה' על בסיס למידה סטטיסטית. נעשה שימוש ביכולות מתקדמות יותר שהתאפשרו בזכות תחילת השימוש בלמידת מכונה, שבמסגרתה נסמכים האלגוריתמים על למידה סטטיסטית מבוססת נתוני עתק. בגל זה, בשונה מהגל הקודם, אין מלמדים את המחשב כללים קבועים אלא מפתחים מודלים סטטיסטיים עבור בעיות שונות, ואז "מאמנים" את האלגוריתמים על דוגמאות רבות, עד שהם מגיעים לרמת הדיוק הרצויה. תוצרי גל זה הם שמאפשרים זיהוי קולי או זיהוי פנים בטלפונים ניידים, או למשל "בוטים" המספקים שירות לקוחות באמצעות התכתבות בצ'טים באינטרנט.

במסגרת גל זה התאפשרה תחילתה של הנהיגה האוטונומית, כפי שהודגמה במסגרת אתגרים בדארפ"א בשנים 2004 ו-2005. עם זאת, לבינה מלאכותית מהדור הזה אין יכולת להבין את הכללים או את הסיבתיים שמאחורי הפעולות שהיא מבצעת, ולכן היא ניתנת להטיה או למניפולציה. עקב התפתחות שנרשמה בשנים האחרונות נוספו שימושים רבים לבינה מלאכותית מהדור הזה, ביניהם מערכות לניתוח או לתרגום טקסט, תוכנות 'עוזר אישי' בטלפון חכם וגם יכולת לשחק משחקים מאתגרים דוגמת משחק האסטרטגיה הסיני 'גו'. הגל השני מתאפיין, לפי דארפ"א, ביכולת לקטלג דברים על פי ניואנסים וביכולת חיזוי, אולם התוצרים חסרי יכולות הקשריות ובעלי יכולות מינימליות של הנמקה הגיונית.

בינה מלאכותית מהגל השני



יכולות סיווג וחיזוי מורכבות

אין יכולת לבצע הקשרים וקיימת יכולת הנמקה הגיונית מינימלית

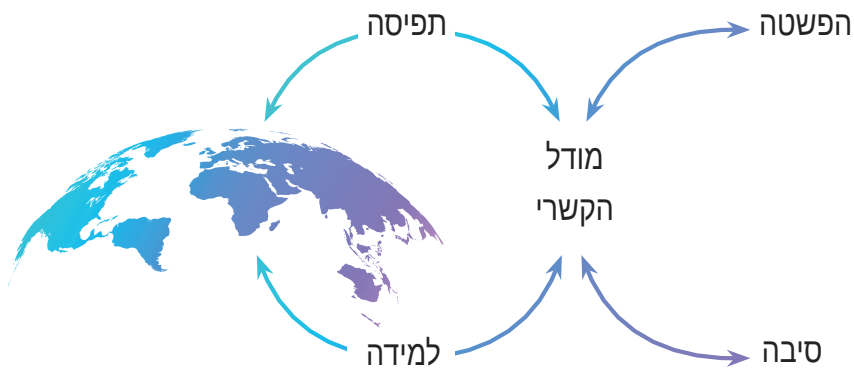
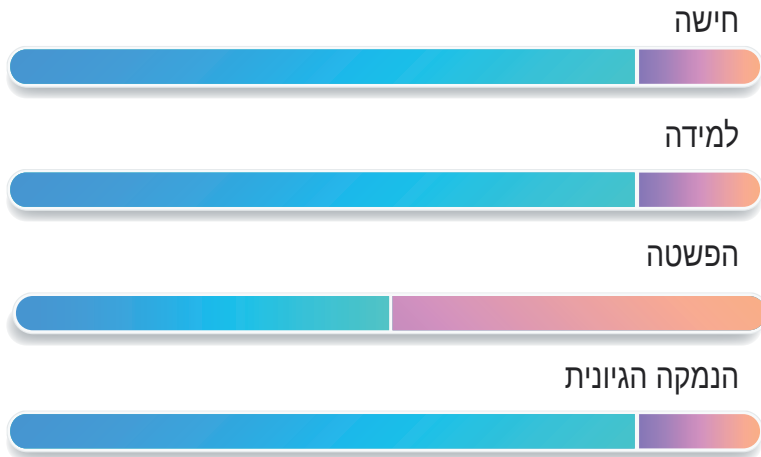
¹⁶ מבוסס על שרטוט מתוך: A DARPA Perspective on Artificial Intelligence

גל שלישי – הסברי: הסתגלות הקשרית

הגל שלישי הוא גל הסברי הנמצא כיום בשלבי פיתוח ומבקש לאפשר 'הסתגלות הקשרית' (Contextual Adaption). האלגוריתמים או המערכות המבוססים על גל זה מנסים לגבש בעצמם מודלים שיסבירו נושאים מסוימים. בדארפ"א מצפים כי מערכות שייבנו סביב מודלים הקשריים ילמדו בעצמן כיצד מודלים שונים צריכים להיות מובנים. כך יתאפשר למערכות להשתמש במודלים עבור מתן הסברים וקבלת החלטות בנוגע לסוגיות ולנושאים שונים. יכולות אלה שונות במידה משמעותית ממרבית האלגוריתמים, הפועלים כיום כ"קופסה שחורה" ויוצרים אתגר בהסברתיות (Explainability) בנוגע לאופן הגעתם למסקנות (נושא שעל אודותיו יורחב בחלק העוסק באתגרים). כך יאפשרו יכולות אלה שימוש במידע באופן מופשט והעברתו צעד קדימה, אולם כיום היכולות של המערכות הללו עדיין מוגבלות.¹⁷ התקווה היא שתוצרי גל זה יהיו "אנושיים" יותר ובין היתר יאפשרו תקשורת בשפה טבעית, יהיו מסוגלים ללמד ולאמן את עצמם (כמו למשל תוכנת 'אלפא-גו' שאימנה את עצמה באלפי משחקי 'גו' נגד עצמה), וכן יכללו יכולות איסוף מידע ממספר מקורות שונים וגיבוש מסקנות מוסברות היטב.¹⁸ על פי הסוכנות, הגל הזה אמור לשפר מאוד את היכולות בתחומים חישה, למידה והנמקה, אך התוצרים בכל זאת ימשיכו להיות בעלי יכולות בינוניות בלבד בתחום ההפשטה.

טכנולוגיות שכבר מתחילות להופיע כיום בהקשר לגל זה הן למשל טכנולוגיות "עוזרות חכמות" המתקדמות ביכולת הסיוע שלהן ביחס לטכנולוגיות של הדור השני, וביניהן סירי ואלקסה.¹⁹ דוגמה נוספת היא תוכנת Google Duplex, המסוגלת להזמין תורים (למשל מקום במספרה או במסעדה) תוך ניהול שיחה קולית קוהרנטית עם נציגי שירות אנושי. מלבד המשימות שתוכנה זו יכולה לבצע באופן אוטונומי, היא גם יודעת לזהות ולאותת למשתמש בה על משימות שהיא אינה מסוגלת לבצע בעצמה.²⁰

בינה מלאכותית מהגל השלישי



²⁴A DARPA Perspective on Artificial Intelligence: מבוסס על שרטוט מתוך:

בנפרד מהחלוקה לגלים, באופן כללי בינה מלאכותית היא מושג המתייחס בכלליות לחומרה או לתוכנה או לשילוב ביניהם, אשר מסוגלים להציג התנהגות שנראית אינטליגנטית. עם זאת, כמעט כל מחקר או מאמר מהשנים האחרונות העוסק בבינה מלאכותית, ובעיקר בבינה מלאכותית וביטחון לאומי, פותח בעובדה שכיום לא קיימת הגדרה אחת מוסכמת למונח בינה מלאכותית.

הבעיה בניסוח הגדרה אחת מוסכמת לבינה מלאכותית נובעת משתי סיבות עיקריות: הראשונה היא הגישות השונות והמגוונות לחקר התחום;²² השנייה היא הקושי הבסיסי להגדיר או להסכים על הגדרת המונח 'בינה' או 'אינטליגנציה', עקב גבולות שטרם נפרצו בחקר מדעי המוח (וגם הפילוסופיה), ולכן מוגבלת היכולת לבחון מושגים אלה ביחס למכונות או להשליך אותם עליהן. חרף הקושי הזה ייבחנו אפשרויות שונות להגדרה, ותוצג הגדרה שתשמש את הדיון בחיבור זה ואת ההמלצות למדיניות שינוסחו על בסיסו.

בינה מלאכותית – הגדרה אופרטיבית

אחת ההגדרות הידועות לבינה מלאכותית, שכבר הוצגה, נוסחה על ידי מרווין לי מינסקי: "לגרום למכונה להתנהג בדרך שהייתה נחשבת לאינטליגנטית לו אדם התנהג כך".²³ יתרונה הוא בהיותה רחבה מספיק ועל כן מסוגלת לכלול רעיונות, שיטות ואמצעים שונים. ואולם, חסרונה המרכזי הוא השימוש במונח 'אינטליגנטי' בהקשר האנושי – מונח שטרם הוגדר ואופיין בצורה חד-ערכית על ידי הדיסציפלינות המדעיות העוסקות בנושא. יתרה מכך, כאשר מעוניינים להגדיר את הדיסציפלינה עבור הביטחון הלאומי ולהמליץ על מדיניות לגביו, יש צורך בהגדרה דיכוטומית יותר שתאפשר החלטה באשר לנכלל בה ולמה שאינו רלוונטי לה.

נטען למשל כי "בינה מלאכותית היא כלי רחב היקף המאפשר לאנשים לחשוב מחדש על האופן שבו נקלט מידע, מנותחים נתונים ומיושמות תובנות הנובעות מכך, כדי לשפר קבלת החלטות". מחברי המחקר הגורס זאת, דארל ווסט וג'ון אלן, סבורים כי אף על פי שאין הגדרה אחידה מוסכמת, נכון להתייחס לבינה מלאכותית כאל "מכונות המגיבות בעקביות לגירוי, בתגובות מסורתיות המיוחסות לבני אדם, בהינתן היכולת האנושית להתבונן, לשפוט ולהתכוון".²⁴ לשיטתם, בינה מלאכותית תלויה בנתונים שניתן לנתחם בזמן אמת ואשר יוכלו להשפיע על בעיות קונקרטיות. תנאי הכרחי לפיתוח בינה מלאכותית מוצלחת, בראייתם, הוא קיומם של נתונים "נגישים לחקירה" בקהילת המחקר.²⁵

לפי שובהנדו שוקלה וויג'יי ג'יסוואל, יישומי בינה מלאכותית "מקבלים החלטות שלרוב דורשות רמת מומחיות אנושית" ועוזרים לאנשים לצפות בעיות או להתמודד עם סוגיות שעולות. כך הם פועלים באופן מכוון, אינטליגנטי ואדפטיבי.²⁶

לאחר הדיון בכמה מן ההגדרות התאורטיות, ראוי לבחון כיצד גורמים העוסקים במחקר ופיתוח או ברגולציה וחקיקה של בינה מלאכותית מגדירים אותה לשם עיסוק פרקטי בה. חרף ההגדרה הכללית של דארפ"א שהוזכרה קודם – "יכולת מתוכנתת לעבד מידע",²⁷ חשוב להבהיר כי לא כל מערכת מחשוב עושה שימוש בבינה מלאכותית. אלגוריתמים של בינה מלאכותית מעוצבים לצורך קבלת החלטות ועושים זאת לרוב באמצעות שימוש בנתונים המוזנים להם בזמן אמת. כאשר הם נמצאים בשימוש על גבי מערכות שונות, לא מדובר במכונות פסיביות שמסוגלות לתגובות מכניות או קבועות מראש בלבד, כפי שהורגלנו בעידן של אוטומציה (כמו דלתות אוטומטיות, או אפילו תוכניות שונות במכונת הכביסה), אלא מדובר במכונות שבעזרת חיישנים, נתונים דיגיטליים ואפילו קלטים (inputs) מרוחקים מסוגלות לשלב את המידע ממקורות שונים, לנתחו מייד ולפעול בהתאם לתובנות הנגזרות מן הנתונים. כך מתאפשרים תחכום ומהירות בקבלת הנתונים, שלא התאפשרו קודם לכן.²⁸

מבחינת הממשל האמריקאי לא קיימת הגדרה אחת רשמית לבינה מלאכותית, וסוכנויות שונות עשויות להגדירה באופן שונה ורלוונטי לצורכיהן. אולם, סדרת חוקים המסדירה את תקציב משרד ההגנה האמריקאי (FY2019 National Defense Authorization Act) מספקת הגדרה של בינה מלאכותית לשם חקיקת סעיף 238, שעוסק במחקר ופיתוח של התחום:²⁹

- כל מערכת מלאכותית שמבצעת משימות, בנסיבות משתנות ולא צפויות, ללא פיקוח אנושי משמעותי, או שביכולתה ללמוד מניסיון ולשפר את ביצועיה כאשר היא נחשפת למאגרי נתונים.
- מערכת מלאכותית המפותחת בתוכנת מחשב, בחומרה פיזית או בהקשר אחר, הפותרת מטלות שדורשות תפיסה, קוגניציה, תכנון, למידה, תקשורת או פעולה פיזית הדומים לאלו של אדם.
- מערכת מלאכותית שמתוכננת לחשוב או לפעול כמו בן אדם, וכוללת ארכיטקטורה קוגניטיבית ורשתות נוירונים.
- קבוצת טכניקות הכוללות למידת מכונה שמתוכננת להעריך מטלה קוגניטיבית.
- מערכת מלאכותית המתוכננת לפעול באופן רציונלי כגון 'תוכנת סוכן אינטליגנטי' (Intelligent Software Agent) או רובוט, שמגשימים מטרות בעזרת שימוש בתפיסה, תכנון, הנמקה, למידה, תקשורת, קבלת החלטות ופעולה.

הגדרה זו מפורטת למדי ואכן מתאימה למטרות חקיקה. היא גם מסייעת, בהשוואה להגדרות אחרות, להחליט אילו תחומי תכנות ומחשוב אינם שייכים לדיסציפלינת הבינה המלאכותית. עם זאת היא ארוכה וטכנית למדי. לאור מטרותו של חיבור זה – להנגיש את הידע על בינה מלאכותית למקבלי החלטות וכן להמליץ על מדיניות

בתחום הביטחון הלאומי – הוא נדרש להגדרה קצרה ופשוטה יותר כמו זו של דארפ"א. הגדרה זו הולמת את מטרות חיבור זה יותר מאשר הגדרתו של מינסקי, למשל, משום שאינה מתייחסת לסוגיית האינטליגנציה האנושית השנויה במחלוקת, ולמעשה היא מאפשרת להכליל במסגרתה מגוון יישומים או שיטות עיבוד המקובלות כיום ואף להשאיר פתח להתפתחויות עתידיות, בלי להכביד בפרטים טכניים שהבנתם דורשת מומחיות. אפילו אם הגדרה זו עשויה לאפשר במסגרתה הכללה של יכולות "נחותות" של מחשוב ועיבוד, כפי שהוסבר בחלוקה לגלים, חלק מן השיטות והתפיסות של הגל הראשון עדיין שימושיות ביותר בתחומים וביישומים שונים, ועל כן יש בכך ערך. ואולם, במקרים שנחוץ למקבלי החלטות לצמצם את רוחב היריעה על מנת לבחון אם פיתוח כלשהו עונה להגדרה של בינה מלאכותית או לא, אפשר להתייחס לבינה מלאכותית כאל אפשרות ליצירת ידע ותובנות שלא הייתה קיימת קודם לכן, מתוך מידע ובאמצעות מכונות ומחשבים. הגדרה זו, המתמקדת ביכולת מתוכנתת לעבד מידע ואשר עלתה בדיוני הוועדה המייעצת למחקר זה, מבחינה בין חלק ניכר מיישומי הבינה המלאכותית לבין יישומי מחשוב באופן כללי, ומאפשרת צמצום ההגדרה הכללית באופן שעדיין יכול לכלול מספר רב של יישומים ולשרת מגוון רב של תחומים, תוך דגש על יצירה של ידע חדש.

ולכן, ההגדרה המנחה חיבור זה היא: **בינה מלאכותית היא שימוש במידע ובמערכות מחשב, המאפשר להציג התנהגות שנראית אינטליגנטית או ליצור ידע ותובנות שלא היו קיימים קודם לכן.** הגדרה זו רחבה מספיק על מנת לכלול טכנולוגיות ויישומים מסוגים שונים, וכן מאפשרת התייחסות נרחבת לצרכים מסוגים שונים על מנת לממש את היכולות הללו. מצד שני, הגדרה זו צרה מספיק כך שלא ייכללו במסגרתה כל תחומי המחשוב באשר הם, אלא רק כאלו שבהם תכונותיה של בינה מלאכותית באות לידי ביטוי. על בסיס הגדרה זו גובשו הפרקים הבאים.