הקמת תשתית חישוב לאומית לבינה מלאכותית

מיפוי דרכי פעולה – מסמך מסכם

ינואר 2026

עריכה: מר **עמרי נצר**, סמנכ"ל טרנספורמציה לענן ו-FinOps, קבוצת יעל

הובלה והכוונה מקצועית: מר **אביתר פרץ**, סגן בכיר לחשב הכללי, דאטה, חדשנות ודיגיטל

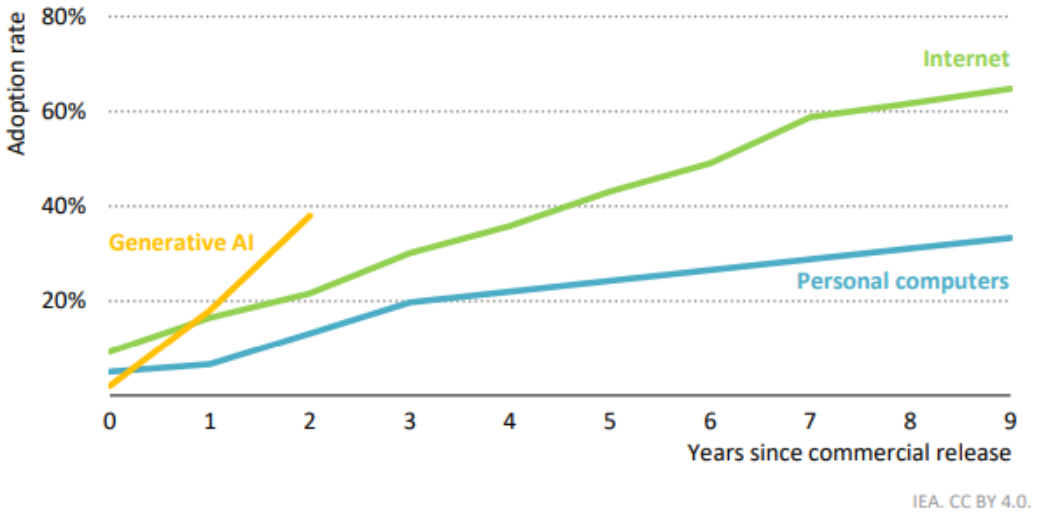
המסמך אושר על ידי:

יהלי רוטנברג
החשב הכללי

ארז אסקל
ראש המטה הלאומי לבינה מלאכותית

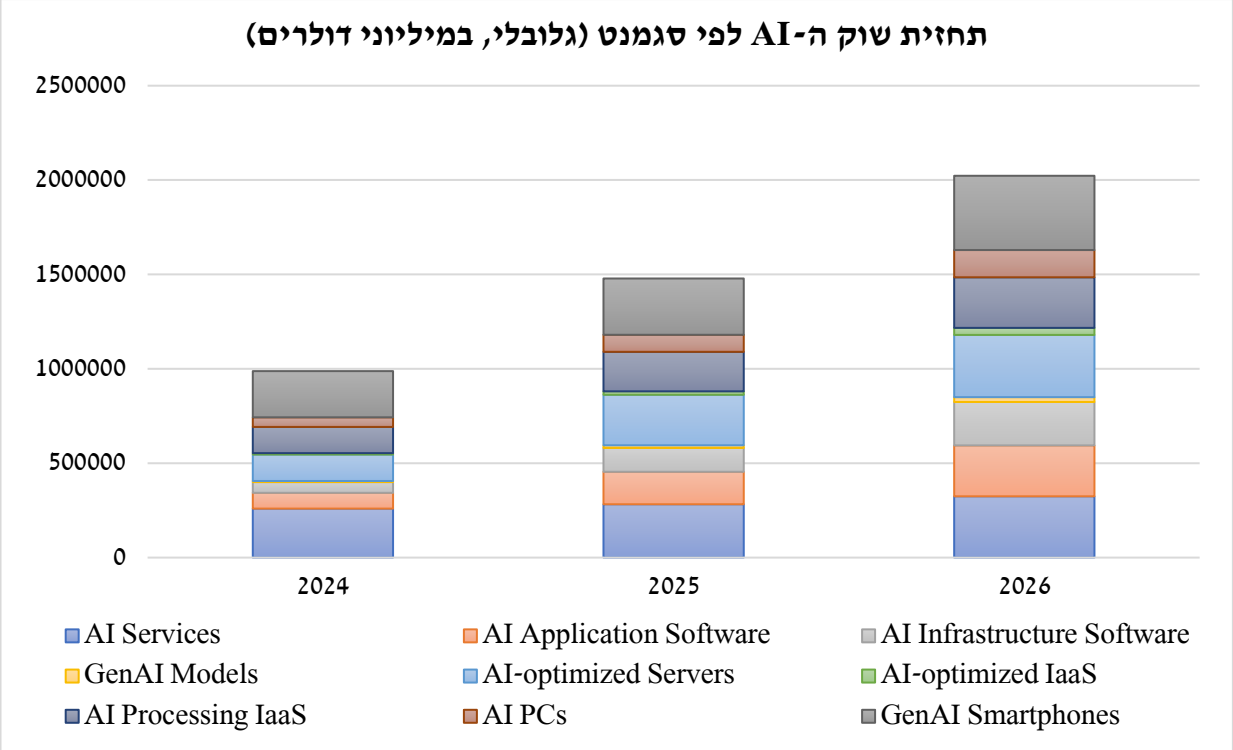
1. מבוא

1.1. בשנים האחרונות חלה התפתחות מואצת בתחום הבינה המלאכותית. הפריצה המשמעותית ביותר של בינה מלאכותית לתודעה הציבורית הרחבה התרחשה בנובמבר 2022 עם השקתו של ChatGPT, אשר בישרה לציבור הרחב על ההזדמנויות הגלומות בבינה מלאכותית, והניעה תהליכי אימוץ חריגים בהיקפם. כ-40% מבתי האב בארה"ב ובריטניה דיווחו כי נעשה בהם שימוש בכלי GenAI זה או אחר, וקצב האימוץ של GenAI גבוה מטכנולוגיות משבשות שקדמו לה:[1]



1.2. ארגונים ומדינות החלו לראות כולם בבינה המלאכותית כוח מהותי אשר יעצב את עתיד העולם. על פי תחזית של קרן המטבע העולמית, הבינה המלאכותית צפויה להשפיע באופן משמעותי על כ-40-60% מהמשרות בעולם. התחזית להשפעה נרחבת נתמכת בניתוחים נוספים, ועל פי McKinsey & Company, PwC ו-IDC, התרומה של בינה מלאכותית לכלכלה העולמית צפויה להסתכם בטריליוני דולרים רבים עד העשור הבא. ארגונים עדיין תרים אחר תועלות כלכליות ממשיות שנובעות מיישומי AI, אולם התחזיות להשפעה האדירה של בינה מלאכותית, כבר מושכות השקעות אדירות לתחום. לפי Gartner, היקף ההוצאה על בינה מלאכותית צפוי לעמוד ב-2026 על יותר מ-2 טריליון דולר:[2]

[1] Laura Cozzi, et al., "Energy and AI," *International Energy Agency (IEA)*, April 10, 2025.
[2] "Forecast Alert: AI in IT Spending, 2Q25," *Gartner*, August 19, 2025.



1.3. מדינת ישראל רואה גם כן בתחום הבינה המלאכותית מרכיב מרכזי בחוסן הלאומי, כבסיס לצמיחה כלכלית-חברתית וכתחום שיש בו כדי לחזק ולשמר את מעמדה של מדינת ישראל בקרב המדינות המובילות בעולם בתחומי הטכנולוגיה בכלל ובתחום הבינה המלאכותית בפרט. מחויבות זו של ממשלת ישראל קיבלה ביטוי בהחלטת ממשלה מס' 3375 מיום 21.09.2025 (להלן: "**החלטת הממשלה**"), אשר במסגרתה הוחלט, בין היתר, על הקמת המטה הלאומי לבינה מלאכותית כיחידת סמך במשרד ראש הממשלה, וזאת בהתאם להמלצת דו"ח הוועדה הציבורית להאצת תחום הבינה המלאכותית (להלן: "**הוועדה הציבורית**"). החלטת הממשלה מתווה את עקרונות היסוד לגיבושה של תכנית להאצת תחום הבינה המלאכותית במשק ולפיתוחן של יכולות מחשוב-על בריבונות ישראלית.

1.4. בהתאם לכך, נערכה עבודת מטה במטרה לבחון חלופות שונות ולסקור מהלכים שנקטו מדינות שונות ביחס להקמתן והפעלתן של תשתיות חישוב לבינה מלאכותית ולמפות את דרכי הפעולה בעניין. מסמך זה נועד לזקק את עיקריה של עבודת המטה שבוצעה ולסייע בקידומה של תכנית ה-AI הלאומית אשר תגובש על ידי המטה הלאומי לבינה מלאכותית ותהווה מצפן למשרדי הממשלה והמשק בכללותו. בתכנית ה-AI הלאומית יוגדרו יעדים, אבני דרך מרכזיות, היקפים ולוחות זמנים למימוש.

2. <u>תשתיות AI – כללי</u>

2.1. ההתפתחויות שחלו בתחום הבינה המלאכותית, ובמיוחד הצורך לעבד, לנתח ולאמן מודלים על כמויות עתק של נתונים, יצרו דרישה מוגברת לכוח חישוב מתאים. על מנת להפיק את מירב הערך מכוח החישוב המוקצה, יש צורך בתשתיות AI מתאימות, אשר מספקות מענה ייעודי, מהיר, יעיל, עמיד ויציב להרצת עומסי עבודה בהיקפים גדולים.

2.2. יש הבחנה בין שני סוגי משימות מרכזיות המבוצעות בעבודת בינה מלאכותית – שלב האימון של מודל, אשר מצריך כוח חישובי רב, זיכרון מהיר בהיקף נרחב ויכולת לבצע סילומיות מקבילית בהיקף משמעותי; היסק (Inferencing) שהוא שלב הביצוע במסגרתו המודל משמש לצורך ביצוע משימות ומילוי הוראות.

2.3. ככל שחולף הזמן, היקף ההוצאה על משימות הסקה צפוי לגדול ולתפוס נתח גדול יותר מהיקף ההוצאה הכולל על משימות בינה מלאכותית, אולם כיום, החלק הארי של הביקוש לתשתיות בינה מלאכותית עודנו נובע מצרכים הנדרשים לאימון מודלי שפה גדולים – LLMs (Large Language Models), כדוגמת המודלים המפותחים על ידי OpenAI, Google, Anthropic וחברות נוספות. לצורך הפיתוח המתמשך והמהיר של מודלים אלו, יש צורך בהשקעת משאבים ניכרים ההולכים וגדלים בקצב מהיר. על פי סקירות שנערכו על ידי Epoch AI, כוח המחשוב הנדרש לצורך אימון מודלים מתקדמים גדל בשיעור מוערך של פי 5 מדי שנה, עלות האימון של מודלי שפה מתקדמים גדלה בשיעור מוערך של פי 3.5 מדי שנה ועלות הרכש של החומרה הנדרשת לאימון מודלים מתקדמים גדלה בשיעור מוערך של פי 2.6 מדי שנה. הביצועים משתפרים גם כן, ו-GPUs הופכים יעילים יותר בשיעור של יותר מ-30% בשנה. על כן, החומרה משתפרת באופן עקבי והופכת יעילה יותר, אולם לאור הצרכים הגדלים נדרשת השקעת משאבים בהיקף הולך וגדל, ובהתאם צרכי האנרגיה לאימון מודלים מתקדמים גדלים גם הם בקצב משוער של פי 2 מדי שנה.

2.4. לאור האתגרים הללו, הקמת תשתיות בינה מלאכותית הפכה לסוגיה מהותית ותעשיית ה-Data Centers מושכת השקעות עתק. לפי ה-Economist, היקף ההוצאה על Data Centers לבדו צפוי לעמוד על יותר מ-3 טריליון דולר בשנת 2028. הגידול העצום בהיקף ה-Data Centers המוקמים כיום בעולם, אינו נטול אתגרים. שניים מהבולטים שבהם נוגעים לצורך הרב של אתרים אלו במים וחשמל. הצריכה של מים ב-Data Centers בארה"ב צפויה לגדול בכ-170% עד 2030 וצריכת האנרגיה צפויה לגדול פי 3 באותה תקופה. למעשה, הגידול בצריכת האנרגיה של Data Centers צפוי להיות כה גדול, עד שמחסור בחשמל צפוי לעכב למעלה מ-30% מפרויקטי ההרחבה של Data Centers בעולם עד 2026. באירלנד, Data Centers צפויים להיות אחראים על כ-28% מצריכת החשמל במדינה עד 2028, וההכבדה שלהם על רשת החשמל הובילה להגבלה מעשית על חיבור Data Centers חדשים באזור דבלין לרשת החשמל.

2.5. למעשה, כפי שעולה, נושא אספקת החשמל הוא מהאתגרים המשמעותיים ביותר בתחום זה. גם בארה"ב קיימים כבר אתגרים ניכרים באספקת החשמל, וחברות אמריקאיות פונות יותר ויותר להקמת תשתיות בפריסה גלובלית, תוך שימת דגש על אספקת אנרגיה זולה וזמינה וסביבה עסקית תומכת.

3. <u>אפשרויות למימוש ושימוש בתשתיות AI</u>

3.1. כאשר בוחנים את דרכי המימוש המרכזיות של תשתיות AI, ניתן לזהות ארבע דרכי פעולה מרכזיות. הראשונה שבהן היא הקמה עצמאית. בדרך זו, ארגונים מקימים בעצמם תשתיות בינה מלאכותית, משימה הכרוכה בהשקעה הונית ותפעולית ניכרת, הן בשלב ההקמה והן לאורך חיי התחזוקה והתפעול של התשתית. דרך פעולה זו מתאימה בעיקר

כאשר מדובר בעומסי עבודה יציבים, וכאשר לגורם המקים את התשתית יש יכולת להקים, לתחזק ולתפעל את הפלטפורמה לאורך זמן ולנצל את יכולותיה באופן מיטבי. אלו אינן מתאימות למרבית הלקוחות.

3.2. דרך פעולה נוספת אשר תפוצתה מתרחבת בשנים האחרונות היא רכישת גישה למאיצים באמצעות Marketplaces, זירות מסחר שבהן משווקים מציעים להשכרה מאיצי AI עם קיבולת פנויה ואשר זמינים לשימוש ב-Data Centers בפריסה מבוזרת. ה-Marketplaces מציעים גישה למשאבים ניתנים להשכרה ובמחיר אטרקטיבי יחסית, אך ביחס להיבטים רבים יכולת הניהול והתפעול שלהם מוגבלת בהשוואה להיצע של ספקים גדולים ומבוססים, והיצע שירותי הפלטפורמה והערך המוסף מוגבל. לכן, הצעת הערך של ה-Marketplaces קורצת כיום בעיקר ללקוחות קטנים, עם צרכים ממוקדים ויכולת שליטה ועצמאות רבה. יצוין, כי יש חברות שפועלות באופן ממוקד בשוק ה-Marketplaces במטרה לטייב את הצעת הערך ולהפוך אותה לאטרקטיבית יותר ללקוחות ארגוניים גדולים.

3.3. החלק הארי של הביקוש לתשתיות AI מתקדמות מסופק כיום על ידי ספקי הענן הגדולים (ובראשם AWS, GCP, Azure ו-OCI). השירותים שספקי הענן מציעים בתחום הבינה המלאכותית הם חלק מסל רחב של שירותים טכנולוגיים המוצעים על ידם בפריסה גלובלית. כיום, ספקי הענן מציעים גישה למגוון מאיצי AI, הן מתוצרת צד ג' (בדגש על Nvidia ו-AMD), והן מתוצרת עצמית, כדוגמת TPUs של Google, Inferentia ו-Trainium של AWS ומאיצי Maia שמפותחים על ידי Microsoft. זאת, לצד שירותי פלטפורמה מתקדמים ומודלי שפה, לרבות מודלי שפה קניינים. עם זאת, ולמרות היקף פעילותם הרחב, ספקי הענן מתקשים כיום לספק את מלוא הביקוש לתשתיות AI, מה שנתן מקום לכניסתם של ספקים חדשים, הקרויים GPU as a Service או Neoclouds, וכן להקמתן של תשתיות AI מקומיות במדינות שונות בעולם.

3.4. ה-Neoclouds מספקים פתרון ייעודי בעולם תשתיות ה-AI, כאשר מטרתם היא לספק מענה ממוקד, יעיל ובמחיר אטרקטיבי, לצורך הרצתם של עומסי עבודה בתחום הבינה המלאכותית על גבי תשתיות מתקדמות. כיום, המאיצים בתשתיות ה-Neoclouds מוצעים לעיתים, לפי מחירי מחירון, בפערי מחיר שיכולים להגיע למאות אחוזים בהשוואה לספקי הענן הגדולים. למרות זאת, המענה של ה-Neoclouds, שהינו ממוקד, אינו תחליפי לפלטפורמות הרחבות של ספקי הענן, כי אם משלים אותן. חלק מספקי הענן הם צרכנים משמעותיים של תשתיות ה-Neoclouds כיום.

4. <u>תשתיות AI וריבונות</u>

4.1. ריבונות דיגיטלית היא נושא מהותי עבור מדינות וארגונים רבים בעולם. המגמה להעדפה של מיזמים דיגיטליים מקומיים החלה עם ה-Sovereign Cloud, אשר "פרויקט נימבוס", שקודם על ידי ממשלת ישראל, היה מבין הסנוניות הראשונות שלו בעולם.

4.2. יצירתה וביסוסה של ריבונות דיגיטלית אינה מלאכת פשוטה, היות שלעיתים בחירה בריבונות באה על חשבון קדמה טכנולוגית, מה שמחייב מדינות לערוך ויתורים, ולהפנות השקעות הולכות וגוברות לצורך פיתוח אקוסיסטם IT מקומי. בעוד שבעבר הנטייה

לעשות זאת הייתה מוגבלת בעיקר לתעשיות הפועלות תחת רגולציה כבדה, ישנן הערכות הסוברות כי מגמה זו תגבר וכי ייתכן שארגונים יתעדפו ריבונות דיגיטלית כקריטריון מרכזי בבחירת תשתיות ענן ו-AI.

4.3. מתוך הדגש הרחב של מדינות על ריבונות דיגיטלית, מתחדד הצורך הממוקד בריבונות AI (Sovereign AI), או המאמץ של מדינות להשקיע בפיתוח שימושי AI במטרה לקדם יעדים לאומיים ומקומיים, תוך צמצום הסיכונים הנוגעים לשימוש ב-AI, הגברת משילות מקומית בשימושים אלו, וחיזוק יכולות ההתאמה של הכלים לצרכים מקומיים. AI ריבוני נתפס כדרך עבור מדינות רבות להטמעת יכולות AI בסביבות ובתהליכי עבודה רגישים, תשתיות קריטיות ועוד, דבר שמחייב גישה ליכולות AI מתקדמות, מותאמות לצרכים מקומיים ואשר ניתנות לבקרה מקומית. כתוצאה מכך, ובמטרה להבטיח תאימות לרגולציות מקומיות ואינטרסים לאומיים, ניכרת מגמה הולכת וגוברת בעולם להקמת פלטפורמות AI מקומיות באמצעות שותפויות ממוקדות.

4.4. כאשר בוחנים יוזמות AI לאומיות, ניתן לזהות מס׳ רכיבים מרכזיים רלוונטיים, כאשר כל מדינה בוחרת לשים דגש על רכיבים שונים בהתאם לצרכיה והעדפותיה:

4.4.1. **הקמת תשתיות חישוב** תחת ריבונות מקומית, תוך ביסוס אספקה תקינה ואמינה של משאבים חיוניים כמו חשמל, תקשורת ומים.

4.4.2. **שילוב ספקי מחשוב** מקומיים המתמחים בתשתיות AI במטרה לייצר יכולת פעולה מקומית בת-קיימא.

4.4.3. **יצירת מסגרת רגולטורית ומדיניות** כלכלית, משפטית וטכנולוגית המסדירה את השימוש בבינה מלאכותית, אבטחת הנתונים ואופן השימוש והשיתוף במידע בתוך המדינה ומחוצה לה.

4.4.4. **פיתוח מודלי שפה** מקומיים תוך שימוש בנתונים רלוונטיים והשימוש בהם.

4.4.5. **יצירת בסיס מו״פ** רחב במדינה, תוך יצירת אקוסיסטם מקומי עשיר לקידום חדשנות ופיתוחים הקשורים בבינה מלאכותית.

4.4.6. **פיתוח ההון האנושי המקומי**, תוך קידום כוח אדם מקומי בעל כישורים ויכולות לפתח, להטמיע, לתחזק ולהשתמש בתשתיות, כלים ויכולות בתחום הבינה המלאכותית.

4.4.7. ביסוס יכולת מקומית הנוגעת ל**תכנון, פיתוח וייצור שבבים** ורכיבי חומרה אחרים בתחום הבינה המלאכותית.

4.5. בחינה גלובלית של הגישות בהן נוקטות מדינות לביסוס יכולות AI לאומיות מגלה כי אין מודל אחד שבו מדינות פועלות, וניכרות גישות שונות שנבחרות לצורך קידומן של יוזמות אלו, כפועל יוצא, בין היתר, של היתרון היחסי של כל מדינה ומאפייניה הכלכליים. להלן סקירה תמציתית של כמה מהתהליכים הבולטים שמקודמים כיום בתחום תשתיות AI במדינות שונות:

ישות/מדינה	עיקרי התהליך
ארה"ב	• המדינה שהניעה את מרוץ ה-AI העולמי. הממשל השיק אומנם תכניות לקידום AI, תוך השקעת משאבים ציבוריים, אולם הגורם המניע המרכזי של התחום (בארה"ב ובעולם), הוא השוק הפרטי במדינה. • החברות המובילות בעולם, רובן ככולן, הן אמריקאיות ומשקיעות סכומים אדירים בתחום. ההשקעות הפרטיות בארה"ב ב-AI מוערכות בכ-109 מיליארד דולר, פי 12 מההשקעות בסין ופי 24 מההשקעות בבריטניה. השקעות אלו לא כוללות את כל פרויקטי התשתיות, ההתחייבויות והתכניות המקודמות בארה"ב. • הממשל האמריקני מקדם את ה-AI Action Plan, תכנית אסטרטגית של הממשל שמטרתה לבסס ולשמר את מעמדה של ארה"ב כמובילת AI עולמית על ידי קידום יוזמות לפיתוח יכולות AI, צמצום חסמים רגולטוריים והקצאת מענקים למימון יוזמות בתחום.
סין	• בסין ניתן לזהות מס' רב של מיזמים, אשר נתמכים על ידי המדינה. חברות סיניות, כמו Alibaba, Baidu ו-Huawei מפתחות מאיצי AI, ומודלי AI סיניים החלו לזכות בהערכה רבה, כאשר מודלים מובילים – DeepSeek, Qwen, Kimi ו-MiniMax – זוכים בציונים גבוהים בהערכות של מודלי AI. • סין משקיעה השקעות ניכרות בהקמת תשתיות AI מקומיות, כולל בשותפות עם ספקים מקומיים, כדוגמת China Mobile, ספק התקשורת הגדול בעולם. • סין אומנם מפגרת אחרי ארה"ב באופן ניכר בהיקף ההשקעות, אולם יש לה יתרון משמעותי בכל הקשור לאספקת החשמל, שהיא קריטית להרצת עומסי עבודה של בינה מלאכותית. כך למשל, ב-2024 לבדה, סין הוסיפה קיבולת חשמל גדולה משמעותית מארה"ב, אירופה והודו יחדיו, והעובדה כי רשת החשמל נשלטת על ידי המדינה, מאפשרת לה לבצע התאמות ולהקים תשתיות במקומות בהם הדבר ישרת את הצרכים שלה באופן מיטבי.
האיחוד האירופי	• האיחוד האירופי פועל ליצירת מסגרת רגולטורית שתטמיע סטנדרטים ואמון בשימוש ב-AI ותבטיח ריבונות מידע ביבשת, וכן לקידום הקמתן של תשתיות AI בתוך תחומי האיחוד האירופי. • בתחילת 2025, חשף האיחוד האירופי תכנית להקמת AI Gigafactories תוך הקצאה של כ-20 מיליארד יורו לצורך כך. תכנית זו היא חלק מתכנית רחבה יותר, InvestAI, שנועדה לקדם השקעות בבינה מלאכותית בהיקף של כ-200 מיליארד יורו. • בחלק ממדינות האיחוד מקודמות יוזמות ייעודיות. כך למשל, בצרפת מקודמת הקמת תשתית בהיקף של כ-1.4GW, ובגרמניה הוכרז על הקמת ענן AI תעשייתי עבור השוק האירופי. בנוסף למיזמים אלו, מס' חברות אירופאיות בולטות פועלות בתחום ומקימות תשתיות AI מקומיות.
הממלכה המאוחדת	• בריטניה נחשבת כיום לשוק ה-AI השלישי בגודלו בעולם. הממשלה מתמקדת כיום בהרחבת תשתיות העיבוד שלה ועידוד תעשייה מקומית מתוך הסתכלות על AI כזרז

ישות/מדינה	עיקרי התהליך
	לצמיחה כלכלית. הדגש המרכזי בפעילות הממשלה בתחום הוא על יצירת סביבה רגולטורית תומכת למגזר העסקי ועידוד השקעות זרות בתחום לצד הקצאת משאבים ציבוריים. • ההשקעות הפרטיות ב-AI בבריטניה נאמדו בשנת 2024 בכ-21 מיליארד דולר, והממשלה הבריטית הקצתה למעלה מ-600 מיליון דולר לתמיכה בחברות AI במדינה, כולל לצורך פיתוח מודלי שפה. בנוסף לכך, בריטניה שואפת להרחיב את בסיס המו"פ ולהגדיל את היקף המחקר בתחום הבינה המלאכותית במדינה פי 20 עד 2030.
הודו	• תכנית ה-AI הריבונית של המדינה נועדה למצב אותה כמובילה גלובלית ובמטרה לפתח מענה ל-Use cases ספציפיים שרלוונטיים למדינה, בדגש על אימון מודלים על מגוון רחב של שפות מדוברות ואשר יפעלו בהתאם לסט ערכים ותרבות מקומי. • הודו מקדמת את תכנית IndiaAI Mission אשר במסגרתה שואפת המדינה להקים כ-35,000 מאיצים, מתוכם כ-18,000 בשלב הראשון. תשתיות אלו מוקמות באמצעות שת"פ עם המגזר הפרטי, והודו מבקשת לעודד שימוש בתשתיות אלו באמצעות סבסוד ממשלתי (כדוגמת ואוצ'רים לשימוש מוזל בפלטפורמות). • כחלק מיוזמה זו, הודו הקצתה יותר מ-150 מיליון דולר לחברות הודיות לצורך פיתוח מודלי שפה מקומיים. אספקת אנרגיה היא אתגר מרכזי עבור המדינה, שכן תשתיות מאיצים צפויים לצרוך כ-3% מאספקת החשמל במדינה.
קוריאה הדרומית	• מקדמת יוזמת AI לאומית במטרה למצב עצמה כאחת משלוש המדינות המובילות בעולם בתחום (AI G3). התכנית כוללת ביסוס מרכז חישוב לאומי לבינה מלאכותית, תוך הגדלת קיבולת מאיצי ה-AI במדינה פי 15, וכן עידוד השקעות פרטיות ב-AI במדינה, לצד הרחבת התמיכה הממשלתית, הגדלת שיעורי החדירה והשימוש ב-AI במדינה (במגזר הפרטי והציבורי) וקידום מסגרת למשילות ושימוש בטוח ב-AI. • דרום קוריאה הכריזה לאחרונה על שותפות עם השוק הפרטי להקמת תשתית מאיצים רחבת-היקף, במסגרתה מתוכננת פריסה של כ-260,000 מאיצים במדינה עד 2030.
איחוד האמירויות הערביות	• המשאבים האדירים העומדים לרשות המדינה, ובמיוחד הון רב וגישה לאנרגיה זולה, משמשים לקידום שותפויות ומיזמים עם חברות רבות. בימים אלו מוקם באיחוד האמירויות הערביות פרויקט Stargate UAE, פרויקט תשתית הבינה המלאכותית הגדול ביותר שמוקם כיום מחוץ לארה"ב, עם קיבולת מתוכננת של כ-5GW.

ישות/מדינה	עיקרי התהליך
	• הפרויקט מוקם בשותפות עם חברות ממשלתיות מקומיות וספקים אמריקאים מובילים. בנוסף לפרויקטים המקומיים, איחוד האמירויות הערביות משקיעה, באמצעות גופים ממשלתיים, בפרויקטים באירופה וארה"ב.
ערב הסעודית	• בדומה לאיחוד האמירויות הערביות, ערב הסעודית משתמשת בשילוב של אנרגיה זולה, קרקע זמינה והשקעות מדינתיות גדולות כדי למשוך חברות זרות. המדינה מקדמת כיום הקמה של תשתיות עם קיבולת מתוכננת של 6.6GW עד 2034. ככלל, פעילותה של ערב הסעודית בתחום מרוכזת תחת הישות הממשלתית Humain.

4.6. המשאבים הכרוכים בהקמת תשתיות AI הם גדולים, ועל אחת כמה וכמה כאשר מדובר בביסוס יכולות AI ריבוני. ההערכה היא שעד 2029 מדינות שמקדמות יוזמות AI ריבוני יידרשו להשקיע לפחות 1% מהתמ"ג שלהן ביוזמות אלו. השקעות אלו באות לידי ביטוי בדמות השקעות ישירות בתשתיות, סובסידיות ממשלתיות לספקים מקומיים המעורבים במיזמים אלו, הזרמות הון לצורך פיתוח מודלים מקומיים ומיזמים לפריסה ושימוש ביכולות AI מקומיות.

5. <u>הקמת תשתיות AI לאומיות – ניתוח מתווים וחלופות</u>

5.1. הוועדה הציבורית הדגישה בעבודתה, באופן שאינו משתמע לשני פנים, את החשיבות של ביסוס יכולות מדינתיות בתחום הבינה המלאכותית. עמדה זו קיבלה ביטוי גם בהחלטת הממשלה. בחינה מקיפה של תשתיות לבינה מלאכותית ושל פרויקטים המקודמים ברחבי העולם, מחדדת את החשיבות של שת"פ בין המגזר הציבורי והפרטי להצלחת מיזמים אלו. חשיבות זו נוגעת לא רק להקמת תשתית המאיצים עצמה, אלא לצורך בניית יכולות מגוונות בכל שרשרת הערך של תחום הבינה המלאכותית, ובכלל זה:

5.1.1. בניית ערוצים לשיתוף פעולה בין הממשלה, חברות פרטיות ומוסדות מחקר לצורך קידום יוזמות מחקר משותפות, שיתוף פעולה טכנולוגי וביסוס פלטפורמות לשיתוף מידע.

5.1.2. קידום שיתוף פעולה למען זיהוי והגדרה של Use Cases קריטיים וקידום יישומי AI לצורך התמודדות עם אתגרים שונים.

5.1.3. בחינה, התאמה ושילוב של סטנדרטים בינלאומיים על מנת להבטיח שהזירה המקומית פועלת באופן תחרותי גם בסביבה הגלובלית.

5.1.4. יצירת פלטפורמות לדו-שיח בשוק המקומי אשר מקדמות חדשנות.

5.1.5. הקמת תשתיות חיוניות באופן שממקסם את היתרונות וממזער את הסיכונים עבור הצדדים המעורבים.

5.2. הנושאים הללו ואחרים הם מרכיבים קריטיים בהסתכלות האסטרטגית שמבקש לקדם המטה הלאומי לבינה מלאכותית על כל תחום הבינה המלאכותית, וזאת באופן שמבטא מחויבות ממשלתית עמוקה לביסוס יכולות בתחום ולבניית שיתופי פעולה עמוקים עם המגזר הפרטי.

5.3. חשוב לציין בהקשר זה, כי מקודמים כיום מיזמים שונים להקמת תשתיות AI במדינת ישראל, כדוגמת אשכול החישוב של חב' Nvidia או התשתית המקומית של חב' Nebius. תשתיות אלו חיוניות ומספקות מענה על צרכים מקומיים מידיים, אולם הן אינן צפויות לספק לבדן מענה על הצורך בזמינות תשתיתית רחבת-היקף בהסתכלות ארוכת טווח. במטרה לייצר קיבולת מקומית שתשרת את צרכי המשק, האקדמיה, המגזר הציבורי וגורמים נוספים, יש צורך בהרחבה משמעותית, וזאת במקביל לפעילות שמטרתה לייצר ביקושי עומק נוספים במשק.

5.4. ישנם מס' נושאים שהם בגדר הנחות יסוד בהתייחס להקמתן של תשתיות בינה מלאכותית:

5.4.1. שכבות החומרה הפיזית הן מרכיבים יסודיים בכל תשתית בינה מלאכותית, אך היא אינה נגמרת בחומרה הפיזית לבדה, ובוודאי שלא במעבדים. ביסוס יכולות מקיפות בתחום מחייבת הסתכלות על שכבות נוספות, ובוודאי שכבות האורקסטרציה, ניטור ובקרה, בדיקות ותיקון תקלות, אבטחה ושירותי פלטפורמה נוספים. יכולות אלו מספקות ערך מוסף משמעותי והן מאפשרות למגוון רחב של גורמים להפיק את מירב הערך מתשתיות אלו.

5.4.2. הקמת תשתית טכנולוגית הפועלת לטווח ארוך היא מהלך שכרוך בסיכונים פיננסיים וטכנולוגיים משמעותיים. לפיכך, ריבוי ספקים הוא חיוני במטרה לצמצם את התלות בספק יחיד, להפחית את הסיכונים הפיננסיים הנוגעים לפעילות, בוודאי בשים לב להשקעות הכספיות העצומות הכרוכות בהקמת תשתיות אלו, ועל מנת להבטיח חדשנות טכנולוגית לאורך זמן וגישה ליכולות מתקדמות ברמת החומרה והתוכנה. כמו כן, חשוב לתת את הדעת בכל הליך על ניידות נתונים ושילוב סטנדרטיים פתוחים, ככל הניתן, במטרה לצמצם Vendor Lock-In.

5.4.3. היקפי ההשקעות המשמעותיים בתשתיות הפיזיות, היכולות הטכנולוגיות המתקדמות הנדרשות להרצת עומסי עבודה מורכבים לאימון מודלים מתקדמים, לצד קצב התפתחות המאיצים, הביאו לקיצור משך חיי המאיצים באופן שעלול לחייב השקעות גדולות ומוקדמות בתחזוקת ציוד, החלפה ושדרוגים. בהתחשב בכך שציוד IT מהווה את החלק הארי מעלויות Data Center, ובוודאי כאשר מדובר בתשתיות חישוב ל-AI, מדובר בצורך בהשקעת הון גדול בהחלפת חלק ניכר מהציוד מדי מס' שנים בודדות. הגם שמאיצים אינם הופכים מיושנים לגמרי בתוך כ-3-4 שנים, ויכולים להמשיך ולשמש במשימות פחות אינטנסיביות, ההשקעה האדירה ברכש תשתיות הופכת את נושא התקצרות תקופת הפחת ומעגלי ההצטיידות לסוגיה מהותית.

5.5. בשים לב לאמור לעיל, בוצעה בחינה של חלופות שונות להקמת תשתיות AI לאומיות. בכלל זה נבחנה אפשרות להקמה עצמאית של תשתיות, מימושן באמצעות שיתוף פעולה מוסדר עם המגזר הפרטי וכן חלופות נוספות:

5.5.1. הקמה עצמאית:

5.5.1.1. מודל של הקמה עצמאית מצוי בצד אחד של ספקטרום האפשרויות, ובמסגרתו המדינה רוכשת ומקימה את התשתית בעצמה, והיא פורסת, מתחזקת, מפעילה ומנהלת את התשתית באופן מלא (ישירות או באמצעות קבלנים), תוך שילוב יכולות קנייניות בתחומי התוכנה, החומרה, פלטפורמות ועוד באמצעות הליכי רכש ציבוריים.

5.5.1.2. במודל זה, המדינה נהנית מרמה גבוהה של ריבונות ושליטה על התשתית, כמו גם מיכולת התאמה לצרכיה. לצד זאת, דרך פעולה זו כרוכה בסיכונים רבים, מטילה על המדינה את מלוא עלות ההקמה, התחזוקה, התפעול והחידוש של התשתית, והמדינה לוקחת על עצמה סיכונים של ניצולת נמוכה והתיישנות טכנולוגית, וכן היא מוגבלת ביכולתה ליהנות מההתמחות, היתרון לגודל והניסיון הנצבר של הספקים בהקמה והפעלה של תשתיות מסוג זה. למרכיבים אלו חשיבות ניכרת היות שהשוק מאופיין בקצבי תכנון ופריסה מהירים של אתרים, במיוחד בשים לב למשך החיים המוגבל של המאיצים, מה שמחייב שימת לב רבה לנושא ה-Time-to-Market.

5.5.2. שיתוף פעולה עם המגזר הפרטי:

5.5.2.1. ברוב רובם של המקרים בהם מוקמות תשתיות AI במדינות שונות, מבוצע הדבר על ידי גורמים פרטיים או במסגרת שיתוף פעולה עם המגזר הפרטי. ביחס לכך, מודלים מעולמות מימון הפרויקטים (Public-Private Partnerships), וכן מודל ממשלתי מכרזי בדומה לזה שיושם במסגרת "פרויקט נימבוס" מאפשרים לנצל את היתרון היחסי של השותפים ולחלוק את הסיכונים והתועלות ביניהם באופן מיטבי, כאשר יש הבדלים בין המודלים השונים המיושמים.

5.5.2.2. במקרה זה, מודלים אלו מאפשרים לרתום את מומחיות השוק והניסיון הנצבר שלו, להעביר את סיכוני ההקמה והתפעול לספק, בעוד שהמדינה לא נושאת בכל העלויות, אלא בהתאם למודל תשלום מוגדר או 'רשת ביטחון' פיננסית שנקבעה. להבדיל ממודל של הקמה עצמאית, במקרה זה המדינה אינה נהנית מריבונות מוחלטת, אך זהו נושא שיש לשקול אותו בשים לב למאפייני התשתית וייעודה, תוך בחינת המענה החוזי שיכול להינתן לצרכי המדינה, והתאמתו להתמודדות עם הסיכונים הפוטנציאליים המזוהים.

5.5.2.3. יצוין, כי יש לשקול את קצב ההתיישנות המהיר של תשתיות טכנולוגיות, ובוודאי תשתיות בינה מלאכותית, כמו גם השילוב של יכולות טכנולוגיות מתקדמות וקנייניות, והצורך להבטיח חדשנות וקדמה טכנולוגית לאורך שנים. בשקלול סוגיות אלו, ובשים לב להתמחות של הספקים, יכול להיות ערך רב בהעדפת חלופה במסגרתה התשתית נותרת בבעלות הספק בתום תקופת הזיכיון/המכרז, וזאת בשים לב להיקף ההשקעות והמשאבים הייעודיים שהקצתה המדינה לאתר, לרבות אנרגיה, תשתיות תומכות, קרקע ועוד.

5.6. שיתוף פעולה עם המגזר הפרטי יכול לבוא לידי ביטוי במספר אופנים, החל מעבודה והתבססות מהותית או בלעדית על ספק מקומי, עבודה ישירה עם שחקן גלובלי מוביל המקים תשתית מקומית באופן ישיר או עבודה עם קונסורציום משלב, אשר במסגרתה הקמת והפעלת התשתית מבוצעת על ידי התאגדות של ספקים מקומיים וגלובליים.

5.7. בנוסף למודלי ההקמה והפעולה האפשריים, יאופיינו מס' מודלים מרכזיים ל'תמיכה' מצד המדינה אשר יכולים לשמש במסגרת פרויקטים מסוג זה, כאשר דרכי תמיכה שונות יכולות להתאים לתרחישי שימוש שונים, לשלבים שונים בתכנית אסטרטגית רחבה להקמת תשתיות חישוב לאומיות ולשילובים שונים:

5.7.1. תמיכה In-kind – אספקת מעטפת תומכת לספקים, בדמות קיצור תהליכי הקצאת קרקע, הבטחת אספקת חשמל, תעדוף חיבור לחשמל וחיבורי תקשורת, קיצור תהליכי אישור רגולטוריים, פישוט וייעול תהליכי ייבוא של ציוד ועוד. תמיכה זו נועדה להגדיל את הוודאות התפעולית והפיננסית של הפרויקט, לצמצם את הסיכון שלוקח על עצמו הספק ולקצר ולפשט תהליכי הקמה ותחזוקה. לתמיכה זו חשיבות גדולה מאוד, היות שהסרת חסמים, צמצום נטל רגולטורי וסיוע מהותי באספקת חשמל, הם קריטיים במטרה לייצר תמריץ למגזר הפרטי להקמת התשתיות, ולאפשר את הקמתן במהירות וביעילות.

5.7.2. תמיכה במימון – מתן תמיכה לספקים המאפשרת להם להוזיל את מחיר ההון, וזאת באמצעות הלוואה ישירה, ערבות מדינה להלוואה או דרך אחרת לגיבוי החוב של הספק באופן מלא או חלקי.

5.7.3. סבסוד – בדרך פעולה זו, המדינה מסבסדת כמות מסוימת מהשימושים של המגזר הפרטי בתשתית. דרך פעולה זאת יכולה לשמש, בין היתר, לקידום מיזמים אשר מקבלים עדיפות גבוהה מצד המדינה.

5.7.4. התחייבות לצריכה – במודל זה, המדינה מתחייבת לרכש קיבולת מסוימת מהספק – רכש קיבולת זה מקנה למדינה גישה מועדפת לתשתית, והמדינה יכולה להשתמש בה באופן ישיר או להקצות אותה לגורם אחר בהתאם למדיניות ותעדופים שיקבעו.

6. סיכום

6.1. ביסוס יכולות AI לאומיות מחייב לא רק תשתית חישוב איכותית, אלא גם בניית שותפויות ויכולות בכל שרשרת הערך של בינה מלאכותית. לישראל יכולות מפותחות במגוון רחב של תחומים, יתרונות יחסיים בקנה מידה עולמי בכל הקשור להון אנושי ולפעילות המגזר הפרטי, אשר מהווים מוקד משיכה להשקעות ורצון גובר של ספקים לבסס פעילות במדינה.

6.2. המטה הלאומי לבינה מלאכותית, מאמץ הסתכלות אסטרטגית רוחבית ומקדם תכנית עבודה רב-שנתית שנועדה לבנות יכולות במגוון תחומים על פני שרשרת הערך של בינה מלאכותית. עבודת המטה, כמו גם הפעילות המתקיימת במשק החופשי, ניצבים בבסיסן של היכולות המפותחות במדינת ישראל בתחום ה-AI.

6.3. מדינות רבות פועלות כיום באופן אקטיבי על מנת להקים תשתיות חישוב ולייצר זמינות מקומית של מאיצים בהיקף נרחב. הקמת תשתיות AI רחבות היקף במדינת ישראל, הכוללות עשרות אלפי מאיצים, היא משימה שאפתנית המחייבת התגברות על חסמים תשתיתיים ויצירת שיתופי פעולה מהירים עם המגזר הפרטי במגוון של תחומים. כפי שניתן לראות, לכל חלופה להקמת תשתית חישוב לאומית יתרונות וחסרונות. לפיכך, ובמטרה להנגיש יכולות AI לאומיות באופן יעיל, תאומץ הסתכלות רב-שנתית תוך התייחסות למס' רבדי פעולה:

6.3.1. **השלמת עבודת מטה שהחלה במטרה לזהות ולצמצם חסמים תשתיתיים ולקצר תהליכים מנהליים**, ובכלל זה בתחום אספקת האנרגיה במטרה להבטיח זמינות מספקת של חשמל לתשתיות שיוקמו, וקיצור ופישוט תהליכים רגולטוריים. החסמים בתחום זה הם מבין הקריטיים ביותר להתמודדות בטווח הקצר.

6.3.2. בטווח הקצר, ועל מנת לבסס תשתיות AI בתוך פרק זמן של השנתיים הקרובות, **תינתן עדיפות למיקוד בהסרת חסמים, תוך בחינת שילובים של מודלי תמיכה במימון מצד המדינה, בהתאם לצרכים, במטרה לקדם הקמת תשתיות אזרחיות במדינת ישראל**.

6.3.3. במקביל, ובמטרה לתת מענה רחב לתרחישי שימוש מקומיים וממשלתיים, תבוצע עבודת עומק לאפיון דרישות המדינה בהיבטי אבטחת מידע, ריבונות ומשילות מקומית ויכולות טכנולוגיות. כחלק מכך, יקודמו **התקשרויות ממשלתיות לשימוש בתשתיות באופן שמסדיר את דרישות אלו. תינתן עדיפות משמעותית למודלים של שיתוף פעולה עם המגזר הפרטי**, באופן שמביא למיצוי מיטבי של יתרונות השותפים. כחלק מההתקשרויות שייווצרו **ישולבו מודלי תמיכה שונים מצד המדינה**, בהתאם לצרכים שיוגדרו ולתרחישי השימוש.

6.3.4. בנוסף לפעילויות להקמת תשתיות חישוב במדינת ישראל, המטה הלאומי לבינה מלאכותית יקדם **שיתופי פעולה אזוריים ובינלאומיים בתחום הבינה המלאכותית**, ובכלל זה גם ביחס לתשתיות AI.

6.3.5. **תיבחן הקצאת משאבים עבור שימושי AI בפלטפורמות ספקי הענן הזוכים במסגרת "פרויקט נימבוס"**. נושא זה יבוצע בתיאום עם ספקי הענן הזוכים, ומטרתו להרחיב את יכולת השימוש של הממשלה במאיצים מתקדמים באזור הישראלי של ספקי הענן הזוכים ולייצר יתירות של תשתיות AI מתקדמות.

7. מקורות מידע נבחרים

"AI Act." *European Commission*. October 23, 2025. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai [accessed: January 20, 2026].

"AI's $4trn Accounting Puzzle." *The Economist*. September 18, 2025.

Anderson, Stephen, Richard Boxshall, and Ali Hosseini. "US$320 Billion by 2030? The Potential Impact of Artificial Intelligence in the Middle East." *PwC*. https://www.pwc.com/m1/en/publications/potential-impact-artificial-intelligence-middle-east.html [accessed: January 20, 2026].

"Artificial Intelligence: Castle in the Cloud." *The Economist*. July 19, 2025.

Barth, Adam, Chhavi Arora, Gayatri Shenai, Jesse Noffsinger, and Pankaj Sachdeva. "The Data Center Balance: How US States can Navigate the Opportunities and Challenges." McKinsey & Company. August 8, 2025.

Bhatt, Shashank. "Govt to Extend IndiaAI Incentives to 8 Homegrown Foundational-Model Teams." *Outlook Business*. September 12, 2025. https://www.outlookbusiness.com/artificial-intelligence/govt-to-extend-indiaai-incentives-to-8-homegrown-foundational-model-teams [accessed: January 20, 2026].

Brown, Andy, Nick Wayth, Romain Debarre, and Wafa Jafri. "Statistical Review of World Energy, 74th Edition." *Energy Institute in collaboration with Kearney and KPMG*. 2025.

Butler, Georgia. "Companies Behind UAE Stargate Offer Additional Details." *DCD*. May 23, 2025. https://www.datacenterdynamics.com/en/news/companies-behind-uae-stargate-offer-additional-details/ [accessed: January 20, 2026].

Castera, Enrique, and Chirag Dekate. "Cool Vendors in AI Specialty Cloud Providers." *Gartner*. August 7, 2025.

Chen, Yawen. "'Neoclouds' Fine Print is a Silver Lining of Sorts." *Reuters*. October 22, 2025. https://www.reuters.com/commentary/breakingviews/neoclouds-fine-print-is-silver-lining-sorts-2025-10-22/ [accessed: January 20, 2026].

Chui, Michael, Eric Hazan, Roger Roberts, et al. "The Economic Potential of Generative AI." *McKinsey & Company*. June 2023.

Courtney, Allie. "NVIDIA and SK Group Build AI Factory to Drive Korea's Manufacturing and Digital Transformation." *Nvidia*. October 30, 2025. https://nvidianews.nvidia.com/news/sk-group-ai-factory [accessed: January 20, 2026].

Cozzi, Laura, Thomas Spencer, Siddharth Singh, et al. "Energy and AI." *International Energy Agency (IEA)*. April 10, 2025.

Duffy, Clare. "The Big Wrinkle in the Multitrillion-Dollar AI Buildout." *CNN*. December 19, 2025. https://edition.cnn.com/2025/12/19/tech/ai-chips-lifecycle-questions [accessed: January 20, 2026].

"EU launches InvestAI Initiative to Mobilise €200 Billion of Investment in Artificial Intelligence." *European Commission*. February 11, 2025. https://digital-strategy.ec.europa.eu/en/news/eu-launches-investai-initiative-mobilise-eu200-billion-investment-artificial-intelligence [accessed: January 20, 2026].

Fioretti, Lapo, Carla La Croce, Andreo Siviero, and Elisabeth Clemmons. "The Global Impact of Artificial Intelligence on the Economy and Jobs: AI will Steer 3.5% of GDP in 2030." *IDC*. August 2024.

"Forecast Alert: AI in IT Spending, 2Q25." *Gartner*. August 19, 2025.

Gupta, Gaurav, Katja Ruud, Ran Xu, and Mukul Saha. "Predicts 2026: AI sovereignty." *Gartner*. October 16, 2025.

Hacker, Markus. "Nvidia and Deutsche Telekom Partner to Advance Germany's Sovereign AI." *Nvidia*. June 13, 2025. https://blogs.nvidia.com/blog/nvidia-deutsche-telekom-germany-sovereign-ai/ [accessed: January 20, 2026].

Hawk, Ryan, Jeoren Van Hoof, Nicki Wakefield, and Allen Webb. "The Leader's Guide to Value in Motion." *PwC*. April 29, 2025.

Jie, Ye-eun. "Korea to Begin Full-Scale AI Buildout in 2026: Minister." *The Korea Herald*. December 15, 2025. https://www.koreaherald.com/article/10636733 [accessed: January 20, 2026].

Jones, Lydia Clougherty, Clementine Valayer, Benjamin Seesel, and Frances Karamouzis. "How Sovereign AI Impacts Enterprises Opportunity and Creates Nonobvious Threat." *Gartner*. October 31, 2025.

Juricic, Louis. "Michael Burry Warns of $176 Billion Depreciation Understatement by Tech Giants." *Yahoo! Finance*. November 10, 2025. https://finance.yahoo.com/news/michael-burry-warns-176-billion-173613512.html [accessed: January 20, 2026].

Katsurashima, Wataru, Naresh Singh, and Evan Zeng. "Competitive Behaviors in the GPU-as-a-Service Market." *Gartner*, April 3, 2025.

Khandabattu, Haritha, and Birgi Tamersoy. "Hype Cycle for Artificial Intelligence, 2025." *Gartner*, June 11, 2025.

Kshirsagar, Mihir. "AI Chip Lifespans: A Note on The Secondary Market." *Princeton University: Center for Information Technology Policy*. December 18, 2025. https://blog.citp.princeton.edu/2025/12/18/ai-chip-lifespans-a-note-on-the-secondary-market/ [accessed: January 20, 2026].

Leswing, Kif. "The Question Everyone in AI is Asking: How Long Before a GPU Depreciates?" *CNBC*. November 14, 2025. https://www.cnbc.com/2025/11/14/ai-gpu-depreciation-coreweave-nvidia-michael-burry.html [accessed: January 20, 2026].

Littman, Michael L., Ifeoma Ahunawa, Guy Berger, et al. "Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report." *Stanford University*. September 2021.

Ma, Jason. "Nvidia and Deutsche Telekom to Build Data Center in Munich, Germany – Report." *DCD*. October 28, 2025. https://www.datacenterdynamics.com/en/news/nvidia-and-deutsche-telekom-to-build-data-center-in-munich-germany-reports/ [accessed: January 20, 2026].

"Machine Learning Trends." *Epoch AI*. January 13, 2025. https://epoch.ai/trends [accessed: January 20, 2026].

Matthews, Karen I. "Agile AI Partnerships: A Public-Private FLEXible and SMART Framework for National Security and Competitive Innovation." *Harvard Kennedy School: Belfer Center for Science and International Affairs*. May 2025.

Mishra, Naveen, Wataru Katsurashima, and Hardeep Singh. "Size the Market Opportunity for AI-Optimized IaaS." *Gartner*. September 30, 2025.

Moss, Sebastian. "Nvidia Claims 3,000 Exaflops of Blackwell Compute is Coming to Europe." *DCD*. June 11, 2025. https://www.datacenterdynamics.com/en/news/nvidia-claims-3000-exaflops-of-blackwell-compute-is-coming-to-europe/ [accessed: January 20, 2026].

Mukhopadhyay, Sushoyan, and Chirag Dekate. "Adopting Physical AI for Real-World Use Cases." *Gartner*. September 16, 2025.

Muppidi, Sri. "OpenAI CEO Sam Altman Forecasts $20 Billion in Annualized Revenue This Year." *The Information*. November 7, 2025. https://www.theinformation.com/briefings/openai-ceo-sam-altman-forecasts-20-billion-annualized-revenue-year [accessed: January 20, 2026].

"National AI Strategy Policy Directions: MSIT Presents Blueprint for Korea's AI Innovation to Achieve AI G3 Status." *Ministry of Science and ICT*. September 26, 2024. https://www.msit.go.kr/eng/bbs/view.do?sCode=eng&mId=4&bbsSeqNo=42&nttSeqNo=1037 [accessed: January 20, 2026].

"Open-Source AI: DeepSeek Aftershocks." *The Economist*. August 23, 2025.

Rahman, Robi. "Leading ML Hardware Becomes 40% More energy-Efficient Each Year." *Epoch AI*. October 23, 2024. https://epoch.ai/data-insights/ml-hardware-energy-efficiency [accessed: January 20, 2026].

Ramamoorthy, Ganesh, and Errol Rasit. "AI's Future Hinges on Efficient Hardware and Distributed Edge Intelligence." *Gartner*. October 23, 2025.

Rogers, Owen. "Neoclouds: A Cost-Effective AI Infrastructure Alternative." *Uptime Institute*. February 26, 2025. https://journal.uptimeinstitute.com/neoclouds-a-cost-effective-ai-infrastructure-alternative/ [accessed: January 20, 2026].

Satariano, Adam, and Paul Mozur. "Saudi Arabia's New Power Play is Exporting A.I. to the world." *The New York Times*. October 28, 2025.

Savov, Vlad. "China's AI Future Looks Like a Long Road to Small Profit." *Bloomberg*. December 3, 2025. https://www.bloomberg.com/news/newsletters/2025-12-03/china-s-ai-future-looks-like-a-long-road-to-small-profit [accessed: January 20, 2026].

Singh, Hardeep. "Unlocking Value Beyond Compute to Expand GPUaaS in AI-Optimized IaaS." *Gartner*. August 27, 2025.

Singh, Hardeep, and Colleen Graham. "Forecast Analysis: AI-Optimized IaaS." *Gartner*. July 15, 2025.

Singh, Virparatap Vikram. "Three National Approaches Towards Sovereign AI." *The International Institute for Strategic Studies (IISS)*. August 28, 2025. https://www.iiss.org/charting-cyberspace/2025/08/three-national-approaches-towards-sovereign-ai/ [accessed: January 20, 2026].

Sevilla, Jaime, and Edu Roldan. "Training Compute of Frontier AI Models Grows by 4-5x per Year." *Epoch AI*. May 28, 2024. https://epoch.ai/blog/training-compute-of-frontier-ai-models-grows-by-4-5x-per-year [accessed: January 20, 2026].

Slok, Torsten. "Hyperscaler Capex spending." *Apollo Global Management*. October 27, 2025.

Smith, David, and Ed Anderson. "Hype Cycle for Cloud Computing, 2025." *Gartner*. June 23, 2025.

Srivathsan, Bhargs, Marc Sorel, Pankaj Sachdeva, et al. "AI Power: Expanding Data Center Capacity to Meet Growing Demand." McKinsey & Company. October 29, 2024. https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/ai-power-expanding-data-center-capacity-to-meet-growing-demand [accessed: January 20, 2026].

Swinhoe, Dan. "TeraWulf and Fluidstack to Develop Texas Data Center in Google-Backed Deal." *DCD*. October 29, 2025. https://www.datacenterdynamics.com/en/news/terawulf-and-fluidstack-to-develop-texas-data-center-in-google-backed-deal/ [accessed: January 20, 2026].

Swinhoe, Dan. "Cipher Mining Secures Fluidstack as a Customer and Google as an Investor." *DCD*. September 25, 2025. https://www.datacenterdynamics.com/en/news/cipher-mining-secures-fluidstack-as-a-customer-and-google-as-an-investor/ [accessed: January 20, 2026].

Swinhoe, Dan. "European Commission Receives 76 Expressions of Interest for AI Gigafactories Initiative." *DCD*. July 2, 2025. https://www.datacenterdynamics.com/en/news/european-commission-receives-76-expressions-of-interest-for-ai-gigafactories-initiative/ [accessed: January 20, 2026].

Tewari, Suranjana, and Osmond Chia. "Nvidia Strikes Bumper AI Deals with Asia Tech Giants." *BBC*. October 31, 2025. https://www.bbc.com/news/articles/cyv862r7l2ro [accessed: January 20, 2026].

"The $3trn Bet on AI." *The Economist*. September 13, 2025.

"U.S. AI Action Plan and Executive Orders on AI: Multi-Industry Global CIO Impacts." *Gartner*. July 30, 2025.

You, Josh, and David Owen. "How Much Power Will Frontier AI Training Demand in 2030?" *Epoch AI*. August 11, 2025. https://epoch.ai/blog/power-demands-of-frontier-ai-training [accessed: January 20, 2026].

Zhang, Tong, Mike Fang, and Ben Yan. "Comparing Generative AI Capabilities Between China and the U.S." *Gartner*. January 10, 2025.