

AI והסיכונים האתיים

הטיות בלתי מודעות, ממה דואגים וממה להיזהר? איך זה קורה? באיזה תחומים?

אודותיי

- יזם הייטק המלווה סטארטאפים העוסקים ב-AI.
- מייסד שותף של Hogen AI, העוסק בציות של מערכות מבוססות AI לרגולציה.
- עד לאחרונה מקים ושותף של חממת החדשנות של חברת Paramount Global. הקים שלושה סטארטאפים בתחומי Entertainment Tech ו-Travel Tech.
- מראשוני תעשיית הסושיאל בישראל, איש שיווק ופרסום עם ניסיון עשיר.
- בעל סטודיו לפיתוח משחקים, מתמחה במשחקים חינוכיים, משחקים לשינוי חברתי וחוויות משחקיות מרחביות.
- פעיל חברתי (הקים את תנועת דרכנו ואת דמוקרטיה).
- מלמד באוניברסיטה מזה עשר בתחומי חדשנות, שיווק ועיצוב משחקים וחוויה.

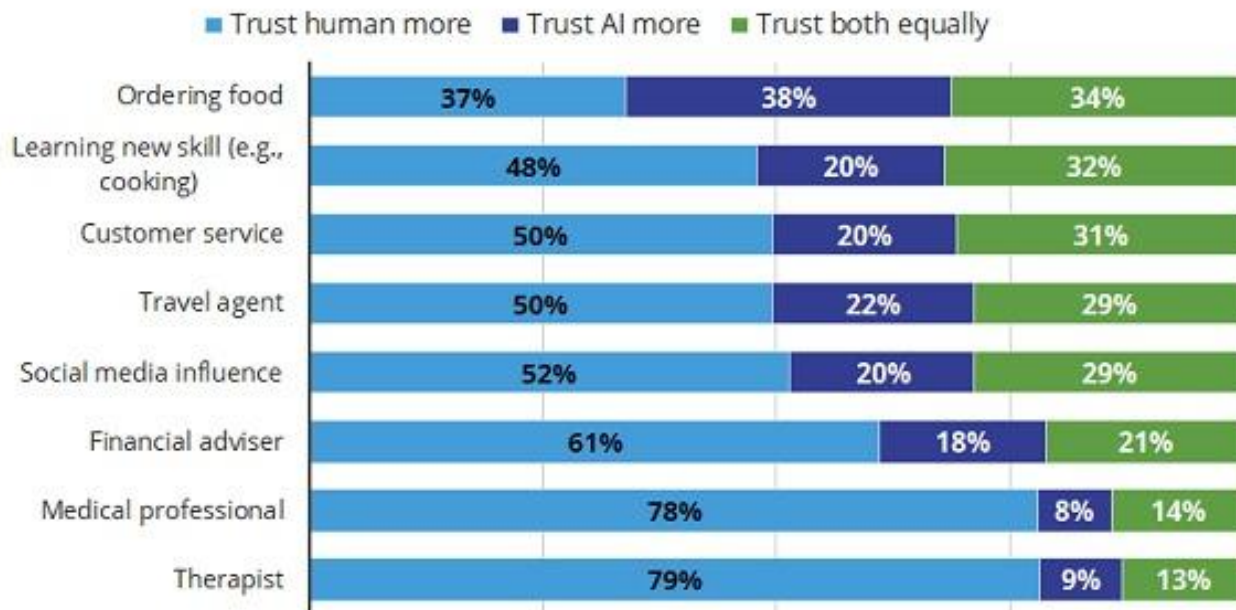


תחומי הסיכון האתיים העיקריים

- הטייה
- שקיפות
- פרטיות
- זכויות יוצרים
- סביבה

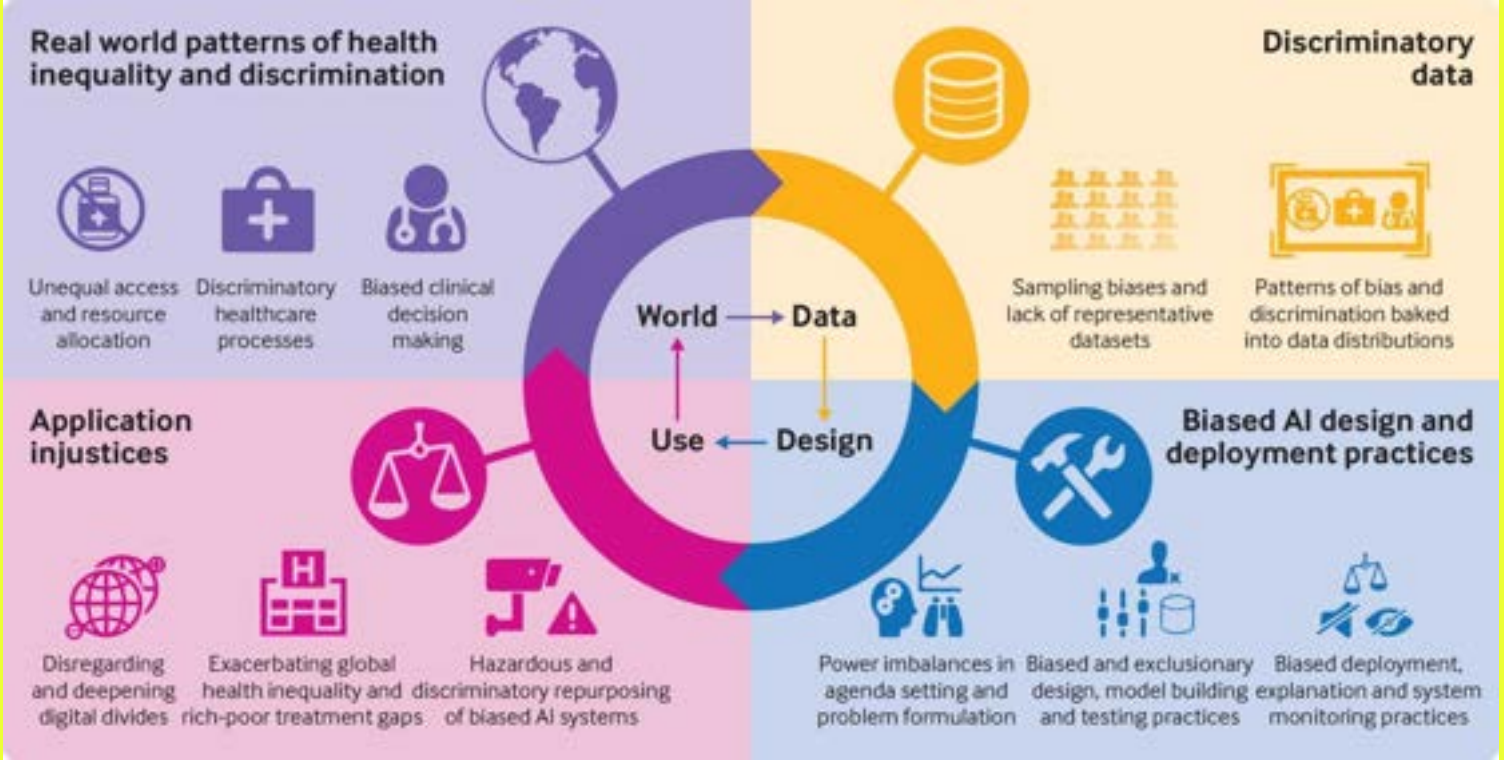
על מי אנחנו סומכים יותר - מכונה או אדם?

What would you trust more for...?

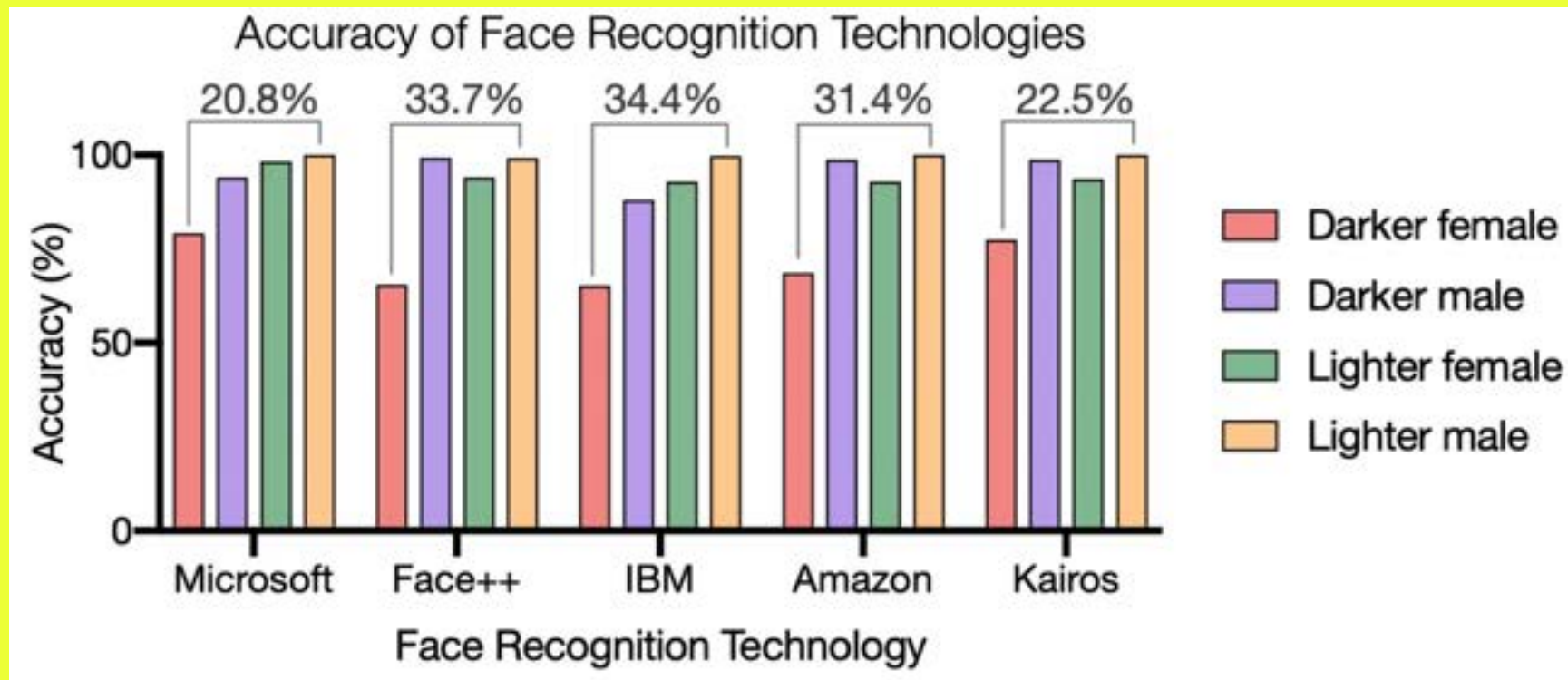


Note: Percentages may not add up to 100% because of rounding
Source: Kearney; base = more than 7,000 global consumers

הטייה



הטייה - מידת זיהוי פנים של לבנים לעומת שחורים (2020)



הטייה - פייסבוק מזהה בוידאו אדם שחור כקוף אדם (2021)



Facebook Apologizes After A.I. Puts 'Primates' Label on Video of Black Men

Facebook called it “an unacceptable error.” The company has struggled with other issues related to race.

שקיפות

- **לפתוח את הקופסא השחורה**
- מה עומד מאחורי התוצאות שאנחנו מקבלים?
 - Explainability - הנגשה של תהליך קבלת ההחלטות על ידי המכונה באופן נהיר וברור..
 - Interpretability - הבנה של הסיבה והתוצאה - מה הקשר בין המידע לבין התוצאות.
- איך מנוהל המידע?
- מה המגבלות של המערכת? מה הסיכונים להטייה בה
- כיצד נבנתה המערכת ומה היו השיקולים המרכזיים מאחוריה?
- איך אנו מפקחים על המערכת עצמה?
- **השקיפות מסייעת לכל ה-Stake Holders להבין כיצד עובדת המערכת, ובכך להגביר את האמון בה**

מכונות טיפשות, ולפעמים גם אנשים

The ChatGPT Lawyer Explains Himself

In a cringe-inducing court hearing, a lawyer who relied on A.I. to craft a motion full of made-up case law said he “did not comprehend” that the chat bot could lead him astray.

ניו יורק טיימס, יוני 2023

אקווריום קרית ים לצלול אל פלאי הים

אקווריום קרית ים הוא אחת האטרקציות המוערכות של העיר ומציע חווית ים עמוקה ייחודית מבלי שתצטרכו לצלול לים באופן פיזי. עם שפע של מינים ימיים מהים התיכון ומחוצה לו, האקווריום המדהים של **רֵית ים קרית ים** מציג מסע חינוכי ומהפנט. תערוכות אינטראקטיביות מאפשרות למבקרים להכיר מקרוב את החיים הימיים, בעוד שאקווריום המנהרה, עם נוף של 360 מעלות, מכניס אתכם ישירות לעולם התת ימי התוסס והמרתק. באקווריום תיהנו מסיוורים מודרכים וסדנאות מיוחדות, מה שמבטיח שגם ילדים וגם מבוגרים יכולים ליהנות ולקבל ידע נרחב על הים.

ישראל היום, ספטמבר 2023

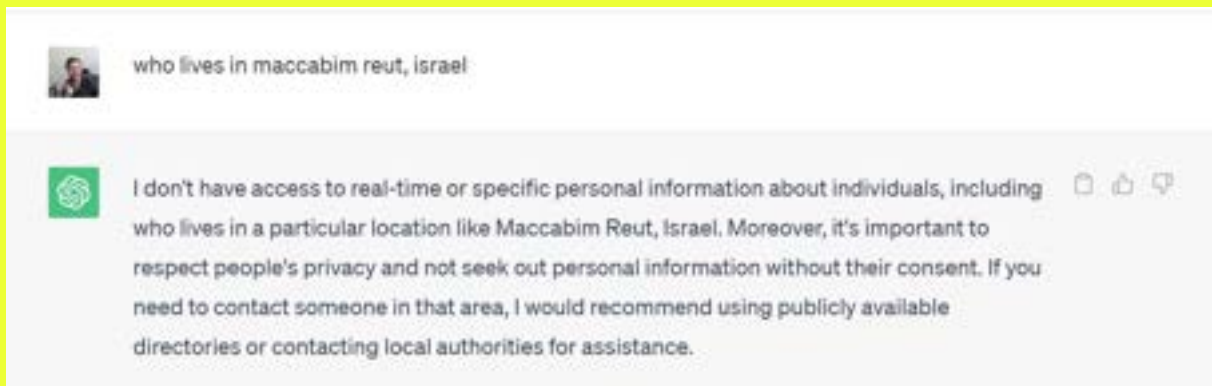
פרטיות

- **המודל עצמו:**

על מה הוא אומן? האם יש בו מידע פרטי של אנשים? פרטי קשר, שמות פרטיים, חשבונות בנק, רשומות רפואיות.

- **מתודת שימוש:**

האם הארגון שאל את עצמו שאלות על פרטיות המשתמשים? האם הוא עושה שימוש בשאלות השונות לאמן את המודל ולשפר את ביצועיו? האם המשתמשים מודעים לכך? האם יש אנונומיזציה של המידע?



Getty Images sues AI art generator Stable Diffusion in the US for copyright infringement



/ Getty Images has filed a case against Stability AI, alleging that the company copied 12 million images to train its AI model 'without permission ... or compensation.'

By James Vincent, a senior reporter who has covered AI, robotics, and more for eight years at The Verge.

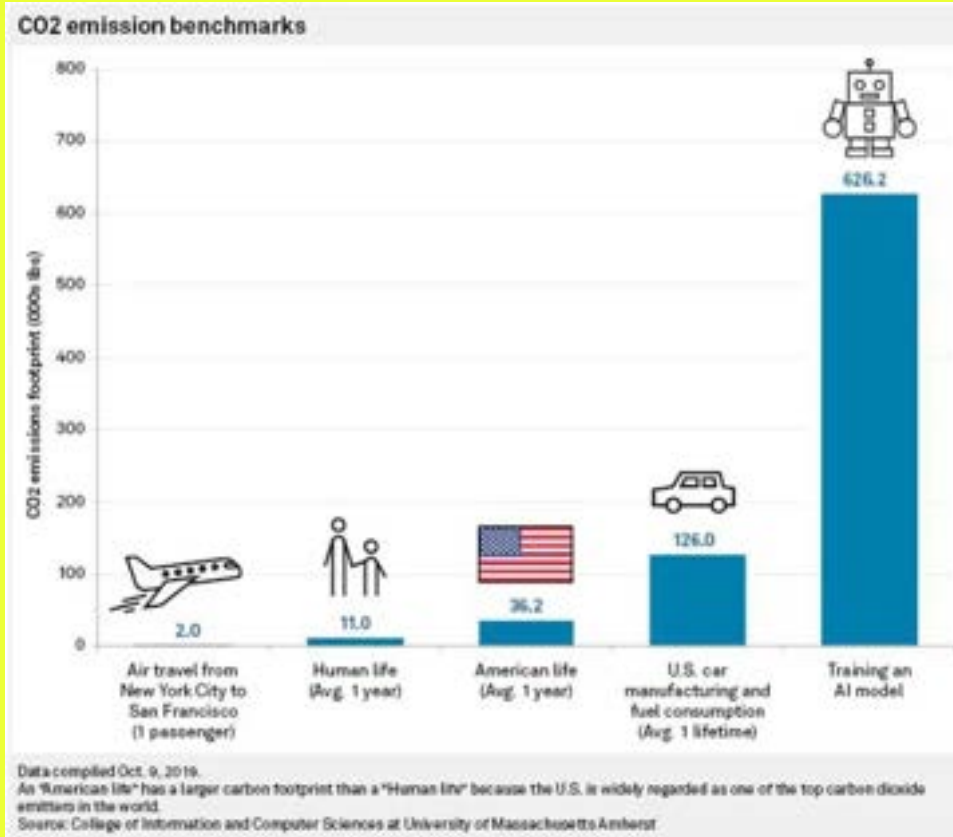
Feb 6, 2023, 6:56 PM GMT+2 | [16 Comments](#) / [16 New](#)

של מי הזכות המוסרית על התוצר?



המחיר הסביבתי

ביוני 2019 פרסמו חוקרים מאוניברסיטת מסצ'וסטס כי כמות האנרגיה הנדרשת לאימון מודל פולטת 626,000 פאונד של פחמן דו חמצני. פי חמש יותר ממה שתפלוט מכונית אמריקאית במהלך חייה.



תודה

nimrod@dweck.co.il

<https://dweck.co.il>