



## כיצד קהילת ה Data Science מסייעת במלחמה נגד קורונה?

# CALLING THE GLOBAL DATA SCIENCE COMMUNITY TO #HACKCORONA

יובל נוח הררי אמר לאחרונה שהאנושות מעולם לא הייתה מוכנה יותר מאשר היום להתמודד עם המגפה. הקידמה הטכנולוגית שמאפיינת את העידן בו אנו חיים לכאורה צריכה להיות לצידנו ולעזור לנו לצלוח את המשבר. יחד עם זאת, המגפה תפסה אותנו לגמרי לא מוכנים.

מי שעוסק בעולם הנתונים לא יכל שלא להיות מתוסכל מהאופן בו החלטות נלקחות בעולם. נדמה שאין מספיק הישענות על נתונים, או מודלים טובים של ניבוי. עם כל ההתקדמות הטכנולוגית האדירה בעולם ה AI ו big data, לא היינו צריכים להיות עכשיו במקום אחר? איך זה קרה? ומדוע אותו AI נודע לשמצה כמי שעתיד למחוק את האנושות – לא מגיע עכשיו ומציל את העולם?

## המשבר הזה חשף את גודל הפער העצום בין ההבטחה הגדולה של עולם הדאטה למציאות, שבה אנחנו לוקחים החלטות ולא מסתמכים על נתונים.

אחת הדוגמאות הבולטות לכך היא היערכות והצטיידות. איך קורה שמישהו ביחידת הרכש לא יודעת להצטייד בציוד מספק? למה אנחנו לא רואים יותר החלטות מבוססות דאטה במשבר הנוכחי?

אפשר בהחלט להתנחם ביכולות הדאטה המצוינות שיש למשרד הבריאות שלנו, שמהווה פורץ דרך בכל עולם הנתונים, ביג דאטה ו Data Science כבר לפני שנים. אך מה קורה בעולם? התחושה היא שמדינות אינן נשענות מספיק על נתונים בקבלת ההחלטות, החל מהבנת תמונת המצב דרך נקיטת פעולות ובוודאי מניעה. וזה בהחלט מתסכל, כי אנחנו נמצאים בנקודת זמן בה היכולות הטכנולוגיות והדיגיטליות קיימות וזמינות ויש כל כך הרבה שניתן לעשות כדי לסייע באמצעות נתונים.

ואכן – בחודש האחרון אנחנו רואים התגייסות אדירה של קהיליית מדעני הנתונים לסייע באתגר. ה קהיליית מאורגנות כמו לדוגמה [Data Natives](#) (כ K80 מדעני נתונים), [Kaggle](#) (גוגל) ודומיהן כבר "על זה" ומריצות האקטונים – דוגמת #HACKCORONA, תחרויות ואתגרים שקשורים למלחמה בנגיף. הרווארד פתחה [קורס ייעודי ל Data Science for Covid19](#) ועוד שלל דוגמאות.



## A fiasco in the making? As the coronavirus pandemic takes hold, we are making decisions without reliable data

By JOHN P.A. IOANNIDIS / MARCH 17, 2020

אז למה בכל זאת קיים הפער הכואב הזה, בין הפוטנציאל והיכולת של עולם הנתונים והמימוש שלו בפועל? באופן מעניין, ניתן להקביל את הסיבות לפער הזה למצב הקיים בארגונים כיום.

### הסיבה העיקרית היא היעדר דאטה. או יותר נכון – היעדר Data governance!

הבעיה היא שהנגיף נכנס לחיינו רק לפני מספר חודשים ספורים ופשוט אין מספיק "היסטוריה". בנוסף, הנתונים שנאספים אינם "סטנדרטיים" משום שמדובר במגפה עולמית וכל אחד מגדיר נתונים טיפה אחרת (לדוגמה, נושא הבדיקות – כל מדינה נוקטת במדיניות שונה. חלק בודקים הרבה, חלק % קטן מהאוכלוסיה, חלק בכלל לא). במצב זה קשה לחזות אחוזי חולי/ תמותה, וכתוצאה מכך גם להיערך ולהצטייד בהתאם. במצב זה, בו אין מספיק נתונים אמנים ואין "אמת אחת" קשה מאוד לבסס מודלים, כפי שאנחנו מכירים היטב מהעולם הארגוני.

החדשות המעודדות הן שיש התקדמות אדירה ומואצת בימים אלה ממש. קהילת ה data science כבר "על זה" ומנסה להתארגן על מקורות מידע טובים. יש כמה יוזמות מעניינות בתחום זה שמנסות להנגיש data sets לכל העולם, כמו לדוגמה [Cord 19](#) - דאטה סט פתוח של Allen Institute המכיל K30 מאמרים ומחקרים בנושא; יוזמות רבות נוספות כמו זו של הבית הלבן בשיתוף עם Kaggle ומיקרוסופט ועוד רבות נוספות שואפות לכנס את קהילת ה AI ומדע הנתונים ולנסות לקדם אתגרים באמצעות אלגוריתמים.

## With launch of COVID-19 data hub, the White House issues a 'call to action' for AI researchers

Taylor Hatmaker @tayhatmaker / 9:15 pm IST • March 16, 2020

Comment

מהם סוגי הפתרונות להם ניתן לצפות משימוש ב Data Science בהקשר של הקורונה?

- בראש ובראשונה, **הבנת הבעיה** (במילים אחרות, איסוף נתונים והנגשתם למי שצריך): היכן ההתפרצויות? מה היקף ההתפרצות? מה אחוז החולי? טכנולוגיות רלוונטיות: ויזואליזציה, דשבורדים, GIS. המחסור העיקרי כאן הוא בתמונת מצב ברורה סוג של Data governance עם "אמת אחת".



- **סיוע באבחון:** אחת הדוגמאות המעניינות כאן היא סיוע באבחון בשלבים מוקדמים, כדוגמת Project Baseline של חברת האחות של גוגל Verily שמציע (כרגע לתושבי קליפורניה) ערכת בדיקה אונליין ל Covid19.



Google Communications

@Google\_Comms

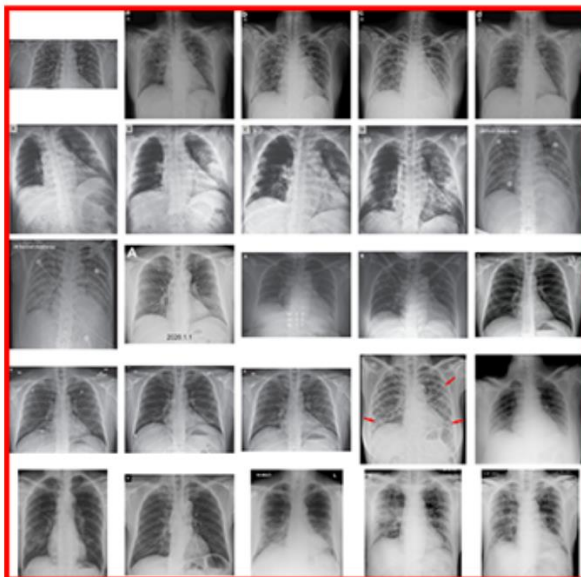
“Google is partnering with the US Government in developing a nationwide website that includes information about COVID-19 symptoms, risk and testing information. (2/6)

6,141 2:44 AM - Mar 15, 2020

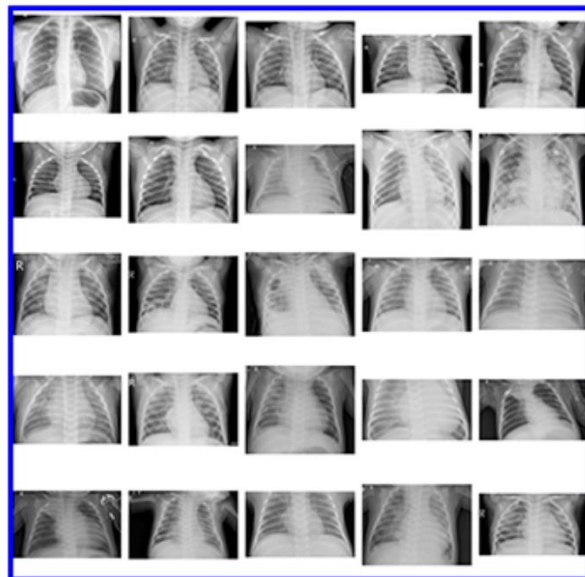
לקריאה נוספת: <https://www.projectbaseline.com/study/covid-19/>

בעולם העיבוד החזותי וספציפית דימות, נעשתה עוד לפני התפרצות הנגיף התקדמות אדירה בפענוח תצלומי רנטגן. כיום data scientists מתחילים להיות מסוגלים לזהות את הנגיף בהתבסס על תמונות דימות בלבד. גם כאן ישנה בעיית זמינות נתונים לאימון המודלים, שמתחילה להשתפר.

COVID-19 +



COVID-19 -





בתמונה – מודל המזהה בדיוק של 90-92% באמצעות Keras, TensorFlow ו-Deep Learning. המודל למד בעצמו, ללא נתונים גיאוגרפיים או דמוגרפיים. המודל מתבסס על data set שפורסם על ידי רופא במונטריאול ב-GitHub ([לינק ל data set כאן](#)) והושגו מול Data Set קיים של Kaggle לצילומי חזה. [עליבאבא פיתחה מודל](#) שמזהה את הנגיף בסריקות CT באחוז דיוק של 96%. המודל אומן על 5000 דוגמאות והמערכת אומצה על ידי 100 בתי חולים בסין וזהו לא המודל היחיד. מודל נוסף פותח באמצעות Deep Learning (תוך שימוש ב K45 סריקות CT) על ידי אוניברסיטת בווהאן – גם שם אחוז הדיוק עמד על 95%.

יש גם התקדמות בשימוש בטכנולוגיות זיהוי פנים לטובת אבחון כפי שאנו רואים בסין [בברוסיה](#), שמסוגלות לבדוק חום גוף וכן להתגבר על בעיות כמו חבישת מסכה לטובת זיהוי פנים.

בהקשר למניעה/מציאת חיסון - בהחלט יש ציפייה ש AI יאיץ את התהליך (שתאורטית לוקח 12 שנה). DeepMind של גוגל הכריזו על שימוש ב Deep Learning על מנת לזרז את הבנת מבנה הנגיף ותהליך קיפול החלבונים באמצעות פרויקט AlphaFold שלה – מערכת שיודעת לנבא מבנה חלבון.

בתוך שלל הדוגמאות לסטארטאפים בעולם AI בתחום מגפות/COVID-19, בולט הסטארטאפ [Blue Dot](#) שכבר היום יכול לצפות מראש היכן המגפה תתפשט (כך אכן עשה בסין לאחר ההתפרצות בווהאן כך שניתן היה להזהיר את הערים הבאות בתור. כמו כן, האלגוריתם מבוסס ML למד את דפוסי התנועה של אנשים ויכל לנבא לאיזה ארצות ייסעו).

דרך נוספת ויצירתית במיוחד לסייע קשורה לכוח מחשובי. לניתוח נתונים בסדר גודל כזה נדרש כוח מחשובי עצום. NVIDIA פנתה לקהיליית הגיימרים בבקשה לתרום את כוחות העיבוד המחשובי שלהם:



## Donate your Computing Power

NVIDIA is asking Gamers to donate their computing power to support folding@home.

### לסיכום

רוב קהילת הנתונים מאמינה של AI יש פוטנציאל רב בסיוע במלחמה נגד קורונה, אבל זה ייקח זמן – המודלים זקוקים לתשתית נתונים טובה, ולזמן אימון. לכן לא תולים בכך הרבה תקוות בעתיד הקרוב.

המפתח ליכולות אנליטיות נעוץ בשכבת נתונים טובה. ככל שמאגרי הנתונים ישתפרו ותהיה "יד מכוונת" (התאגדות של ממשלות לצורך העניין) נוכל לראות יותר פריצות דרך בניסיון לצלוח את המשבר הזה.

כך או כך, אין ספק של AI יש הזדמנות פז להפוך את המוניטין הרע שקיבל מזה שרוצה להשמיד את האנושות לזה שיעזור להציל אותה.

בברכת בריאות טובה לכולם וחזרה במהרה לשגרה "משעממת".

צוות המחקר של STKI.