

פרויקט גמר – חקר ספרות

Voice Speech Dictation

**מאת:**  
אופיר בן זקן 208839712   
מתן טל 201492881

**מנחת הפרויקט:**  
ד"ר תמר שרוט

# 

תוכן עניינים

[רקע תיאורטי - מה זה Voice Speech Dictation ? 3](#_Toc85574071)

[במה המערכת שלנו יכולה להועיל למשתמש ? 3](#_Toc85574072)

[מטרת הפרויקט 3](#_Toc85574073)

[Machine learning 4](#_Toc85574074)

[האלגוריתמים בהם נשתמש 5](#_Toc85574075)

[המודלים במערכת 5](#_Toc85574076)

[המצב הקיים בשוק 6](#_Toc85574077)

[References 7](#_Toc85574078)

# רקע תיאורטי - מה זה Voice Speech Dictation ?

מערכת זיהוי קול/דיבור היא מערכת ממוחשבת המסוגלת לפענח [דיבור](https://he.wikipedia.org/wiki/%D7%93%D7%99%D7%91%D7%95%D7%A8) אנושי. הקלט למערכת מגיע מ[מיקרופון](https://he.wikipedia.org/wiki/%D7%9E%D7%99%D7%A7%D7%A8%D7%95%D7%A4%D7%95%D7%9F) שאליו מדבר המשתמש, משיחת [טלפון](https://he.wikipedia.org/wiki/%D7%98%D7%9C%D7%A4%D7%95%D7%9F) הנקלטת במערכת וכדומה. נהוג להבחין בין זיהוי דיבור שמטרתו לזהות את תוכן הדיבור (לצורך ביצוע פקודה או רישום טקסט) לזיהוי קול שמטרתו לזהות דובר לא ידוע, לאמת זהות דובר ידוע או לאפיין את הדיבור שלו (למשל, בעבר נעשה ניסיון להעריך האם אדם מסוים דובר שקר או אמת לפי הקול שלו). [1] המטרה שלנו בפרויקט היא לייצר אפליקציה אשר יודעת להאזין לקולו של המשתמש, לזהות את המילים שיוצאות לו מהפה, להמיר אותן לטקסט עם כמה שפחות שגיאות כתיב ולשמור את זה כקובץ. זו תהיה מערכת שתתבסס על machine learning, מתחום הבינה המלאכותית שתאפשר למחשב ללמוד תוך כדי עבודה מבלי להיות מתוכנת באופן ספציפי. מטרת למידת המכונה יכולה להיות מידול, חיזוי או גילוי של עובדות לגבי העולם האמיתי. המערכת שנרצה לפתח "זיהוי דיבור קולי ותמלול " תשתמש במערכת למידת מכונה כדי ללמוד, בהינתן אותות קוליים כלשהם, מהי ההברה שיצרה אותם. [2] כיום קיימות המון מערכות שמבצעות את הפעולה ואך יודעות לתמלל בשפות שונות. למרות כל מה שקיים בשוק אנו נרצה לבנות אפליקציה משלנו המתבססת על קיימות אך נוסיף לה פיצ'רים נוספים ונשדרג אותה על מנת ליצור משהו שעוד לא נראה.

# במה המערכת שלנו יכולה להועיל למשתמש ?

המשתמש יוכל לחסוך לעצמו זמן, כאשר בימינו זהו מצרך הכרחי .זמן ההקלדה של המילים יחסך לו ומה שיצטרך לעשות זה רק ללחוץ על כפתור ולדבר, כמובן לא במהירות גבוהה מדי שהמערכת תוכל להבין את המילים היוצאות מפיו. מעבר לכך המשתמש יוכל להשתמש במערכת לאו דווקא במחשב אלא גם smartphone, שעון חכם, iPad וכדומה, מה שהופך את השימוש במערכת לנגיש יותר גם מחוץ לבית.

# מטרת הפרויקט

מטרתנו העיקרית היא בעצם ליצור מערכת זיהוי דיבור קולי ותמלול ולשדרג אותה בפיצ'רים נוספים שלא קיימים היום בשאר המערכות הדומות שקיימות, בנוסף נרצה שהתמלול ישמר למשתמש כקובץ כאשר הוא יסיים. נרצה שהמערכת תהיה נגישה לכלל האוכלוסייה ותשפר את איכות חייהם בחסכון בזמן בעבודה במחשבים ובמכשירים חכמים כמו למשל הקלדה במנועי חיפוש, כתיבת מסמכים ארוכים, כתיבת הודעה ועוד. בנוסף כל משתמש יתבקש לבצע לפני השימוש הגדרה של מילות מפתח המתאימות ליצירת פורמט אישי על ידי זיהוי המילים המוגדרות על ידיו למשל המילה end תומר על ידי המערכת לסימן הפיסוק "נקודה" (.).

# Machine learning

למידת מכונה (Machine Learning או בקיצור ML) היא תת-תחום במדעי המחשב והבינה המלאכותית (AI) העוסק במגוון משימות חישוביות בהן התכנות הקלאסי לא ישים. המאפיין המרכזי של למידת מכונה הוא בפיתוח אלגוריתמים שאינם מבוססים על סט חוקים מוגדר מראש, אלא לומדים מתוך מצבור דוגמאות. המטרה המרכזית של למידת מכונה היא טיפול ממוחשב בנתונים, על ידי מידול, חיזוי או גילוי עובדות מן העולם האמיתי, עבור בעיות שלא ניתן לכתוב להן תכנת מחשב "קלאסית". כלומר, אם בתכנות הקלאסי כתיבת התוכנה מתבססת על לוגיקה ברורה שמורכבת מהגדרת תנאים מסוימים (בהם לולאות, דרכי פעולה במצבים שונים ועוד), אזי למידה מכונה (Machine Learning) מאופיינת [באלגוריתמים](https://www.bdo.co.il/he-il/%D7%91%D7%9C%D7%95%D7%92%D7%99%D7%9D/bdo-tech-hub/%D7%9E%D7%A8%D7%A5-2019/%D7%90%D7%9C%D7%92%D7%95%D7%A8%D7%99%D7%AA%D7%9D-(algorithm))שמטרתם לאפשר למחשב ללמוד את התנאים ולהסיק מידע באופן עצמאי מתוך כמות גדולה של דוגמאות. יכולת זו יעילה מאוד בסיטואציות מורכבות בהן קשה להגדיר מראש סט חוקים או תנאים- מנהיגה אוטונומית, דרך אבחון גידולים סרטניים, ועד לחיזוי העדפות משתמש. [11] למידת מכונה כוללת אוסף ידוע של כמה עשרות אלגוריתמים ניתן לסווג אלגוריתמים אלו על פי מספר תבחינים :**1. עצלנים (lazy ) מול חרוצים (eager )–** אלגוריתם עצלן לא מבצע חישוב עד שלא מתבקש לענות על שאלה, כאשר נשאלת שאלה האלגוריתם אוסף נתונים רלוונטיים, מבצע חישוב ונותן תשובה. לעומת זאת האלגוריתמים החרוצים משתמשים בנתוני הלימוד על מנת לבנות מודל מתמטי אשר מהווה בסיס לפתרון הבעיה. כאשר האלגוריתם מתבקש לענות על שאלה הוא משתמש בנתוני השאלה על מנת לפתור את המודל. [2] **2. מקומי (local) מול גלובלי (global)-**אלגוריתם מקומי הוא אלגוריתם שמחשב פתרון שתקף רק לסביבה שבה נשאלה השאלה. אלגוריתם גלובלי לעומתו מחשב מודל שאמור להיות תקף בכל מרחב הנתונים הרלוונטי.[2] **להלן מספר אלגוריתמים נפוצים:**

* אלגוריתם שכן קרוב
* עץ החלטה לומד
* רגרסיה מקומית
* רשת עצבית מלאכותית
* רשתות ביסיאניות
* אשכולות מבוססי מרחק
* אלגוריתם ציפייה – מקסום
* אלגוריתם גנטי
* אלגוריתם של לימוד מותקן (boosting ,bagging)
* חוקים אסוציאטיביים
* PSO- אופטימיזציה מבוססת תבונת נחיל
* ACO – אופטימיזציה מבוססת תבונת נמלים
* מכונת וקטורים תומכים

# האלגוריתמים בהם נשתמש

כמו שכבר הזכרנו קודם לתחום של למידת המכונה יש אלגוריתמים רבים שנעשה בהם שימוש על מנת להגיע ללימוד הטוב ביותר למכונה ושתעבוד בצורה המדויקת ביותר. בפרויקט שלנו אנחנו עובדים עם משהו קיים ומשפרים אותו ולכן נשתמש ב-API של גוגל, בשל כך אנחנו לא יכולים לדעת בדיוק באיזה אלגוריתמים נעשה שימוש . באופן כללי אנו יודעים שיש שימוש ב-Machine Learning (ML) וב- Artificial intelligence (AI). מכאן אנו סבורים שהם השתמשו באלגוריתמים כמו: Natural Language Processing – NLP שתפקידו לעסוק בבעיות הקשורות לעיבוד ומניפולציה על שפה טבעית והבנה של שפה טבעית על מנת לגרום למחשבים "להבין" דברים שנאמרים נכתבים בשפות אנושיות. [10] ואלגוריתמים דומים נוספים..

# המודלים במערכת

**מודל משתמש** – מודל המשתמש יכיל בתוכו את המילים השמורות שיתבקש להגדיר מראש לפני תחילת ההקלטה, מילים אלו יהוו עבורו מעין מילון כאשר ירצה להשתמש בסימני הפיסוק השונים בקובץ שלו. המשתמש לא יוכל להתחיל את השימוש באפליקציה ללא הגדרת מילים אלו. למשל: המשתמש יגדיר שכאשר המערכת תשמע את המילה "new line " באנגלית היא תשמור אותה במילון של אותו משתמש וכאשר היא תשמע את המילה "new line " תתבצע ירידת שורה בקובץ שיתקבל בסיום ההקלטה וכך הלאה.  
  
**מודל מכונה** – במודל המכונה לאחר שהמשתמש יקליט את הקטע אותו ירצה להפוך לטקסט ההקלטה שנקלטה במערכת תיכנס אל תוך המכונה ושם תעבור תהליך עיבוד והמכונה תמיר אותה לקובץ בפורמט שיבחר על ידי המשתמש (pdf ,word).

# המצב הקיים בשוק

כיום קיימות המון אפליקציות עם תפקיד זהה לזו שאנחנו רוצים לפתח . כולן מבצעות את אותה פעולה, זיהוי קולי ותמלול אך לכל אחת יתרונות וחסרונות משלה נרחיב על כמה מהן:  
  
**1. Dragon Professional**– נקראת גם dragon for PC, אפליקציה שיכולה לשמש כל אחד מתלמידים/סטודנטים ועד לאנשים מקצועיים ולעסקים. האפליקציה יכולה לעזור במשימות יום יומיות כמו: כתיבת שיעורי בית, שליחת מיילים ואפילו גלישה באינטרנט בעזרת כך שהמשתמש מדבר והיא כותבת. ידועה בשל יכולות ההכתבה והתמלול המהירות שלה ומסייעת לאנשים עובדים ביצירת מסמכים.[4]  
  
**2.Google Now –** תוכנה של מנוע החיפוש של גוגל (google search), זמינה למכשירי android ו-IOS . התוכנה משולבת היטב עם מערכת ההפעלה של android כך שניתן להשתמש בה לביצוע של כל פונקציונליות הקיימת בה, זמינה גם במחשבים אישיים כחלק מדפדפן האינטרנט google chrome.   
במכשירי אנדרואיד ניתן להשתמש ב-google now לקבלת שיחות, שליחת הודעות טקסט ופתיחה וסגירה של האפליקציה. עבור מכשירי IOS ניתן להשתמש בה לצורך חיפוש פונקציונלי. הפעלת התכונה מתבצעת על ידי הקלקה או נגיעה בסמל הקטן של המיקרופון, בצד חלון החיפוש או, בטלפונים ניידים עם גרסת "[אנדרואיד](https://he.wikipedia.org/wiki/%D7%90%D7%A0%D7%93%D7%A8%D7%95%D7%90%D7%99%D7%93_(%D7%9E%D7%A2%D7%A8%D7%9B%D7%AA_%D7%94%D7%A4%D7%A2%D7%9C%D7%94)) " מתאימה, ישירות על ידי הפקודה הקולית: OK Google כאשר התכונה מופעלת יופיע חיווי מתאים ואז היישומון מאזין לשאלות של המשתמש. תכונה זו אמורה להיות שימושית במיוחד במצבים בהם הידיים תפוסות, כמו למשל בעת בישול במטבח, עבודה עם כלים מכניים או אפילו, בזמן נהיגה.[5]

**3.Siri –** סירי היא מערכת בינה מלאכותית אוניברסלית מסוג עוזרת אישית ווירטואלית למכשירי אפל הנתמכת על ידי 21 שפות שונות. התוכנה מותקנת מראש במכשירי אפל ומגיבה בקולה של סירי. המערכת מעבדת [שפה טבעית](https://he.wikipedia.org/wiki/%D7%A9%D7%A4%D7%94_%D7%98%D7%91%D7%A2%D7%99%D7%AA) [מדוברת](https://he.wikipedia.org/wiki/%D7%93%D7%99%D7%91%D7%95%D7%A8), כדי לענות על שאלות, לתת המלצות, לבצע פעולות בהתאם לבקשות המשתמש ולענות על שאלות באמצעות משאבי המידע [וולפרם אלפא](https://he.wikipedia.org/wiki/%D7%95%D7%95%D7%9C%D7%A4%D7%A8%D7%9D_%D7%90%D7%9C%D7%A4%D7%90) ו-[וויקיפדיה](https://he.wikipedia.org/wiki/%D7%95%D7%99%D7%A7%D7%99%D7%A4%D7%93%D7%99%D7%94). מאפייני המערכת כוללים הסתגלות להעדפותיו של המשתמש לאורך זמן והצגת תוצאות אישיות, בנוסף ליכולת להזמין מונית או אפילו להזמין מקום במסעדה לארוחת ערב . סירי יכולה לבצע שיחות, לשלוח הודעה, לבצע חיפוש באינטרנט, להגדיר אזעקות, טיימרים ותזכורות. [6]

**4. Amazon Lex –** שירות לבניית ממשקי שיחה לכל יישום באמצעות קול וטקסט. התוכנה מספקת את הפונקציונליות המתקדמת של למידה עמוקה של זיהוי דיבור אוטומטי להמרת דיבור לטקסט והבנת שפה טבעית כדי לזהות את כוונת הטקסט, לאפשר למשתמש לבנות אפליקציות עם חוויות משתמש טובה ושיחה אמיתית. [7] האפליקציה ידועה בכינוייה אלכסה, אלכסה היא סייעת וירטואלית שמפותחת על ידי אמזון. אלכסה מסוגלת לנהל שיחות, להשמיע מוזיקה, להכין רשימות, [להגדיר התראות](https://he.wikipedia.org/wiki/%D7%A9%D7%A2%D7%95%D7%9F_%D7%9E%D7%A2%D7%95%D7%A8%D7%A8), להזרים פודקאסטים, להשמיע ספרים קוליים, להציג את תחזית מזג האוויר, מצב התנועה, ספורט, ועוד מידע בזמן אמת, כגון [חדשות](https://he.wikipedia.org/wiki/%D7%97%D7%93%D7%A9%D7%95%D7%AA_(%D7%AA%D7%A7%D7%A9%D7%95%D7%A8%D7%AA)). אלכסה יכולה גם לשלוט במספר [מכשירים חכמים](https://he.wikipedia.org/w/index.php?title=%D7%9E%D7%9B%D7%A9%D7%99%D7%A8%D7%99%D7%9D_%D7%97%D7%9B%D7%9E%D7%99%D7%9D&action=edit&redlink=1) באמצעות מערכת [בית חכם](https://he.wikipedia.org/wiki/%D7%91%D7%99%D7%AA_%D7%97%D7%9B%D7%9D). רוב המכשירים שעליהם פועלת אלכסה מאפשרים למשתמש להפעיל את הסייעת הקולית על ידי אמירת מילה (למשל :"אלכסה").[8]

# References

1. [מערכת זיהוי דיבור – ויקיפדיה (wikipedia.org)](https://he.wikipedia.org/wiki/%D7%9E%D7%A2%D7%A8%D7%9B%D7%AA_%D7%96%D7%99%D7%94%D7%95%D7%99_%D7%93%D7%99%D7%91%D7%95%D7%A8)
2. [למידת מכונה – ויקיפדיה (wikipedia.org)](https://he.wikipedia.org/wiki/%D7%9C%D7%9E%D7%99%D7%93%D7%AA_%D7%9E%D7%9B%D7%95%D7%A0%D7%94#%D7%9E%D7%98%D7%A8%D7%95%D7%AA_%D7%95%D7%A9%D7%99%D7%9E%D7%95%D7%A9%D7%99%D7%9D)
3. [10 Best Voice Recognition Software (Speech Recognition in 2021) (softwaretestinghelp.com)](https://www.softwaretestinghelp.com/voice-recognition-software/)
4. [Buy Dragon Professional Individual - Speech Recognition | Nuance](https://shop.nuance.co.uk/store/nuanceeu/en_GB/Content/pbPage.dragon-professional-individual?currency=EUR&pgmid=95401100&campaignid=1476392048&adgroupid=60832199247&targetid=kwd-329879188924&matchtype=e&device=c&gclid=Cj0KCQjwkbuKBhDRARIsAALysV4QjkDQryaYMDD6LvNkDAazSmm9BzFlgYfI4IxKM8QQ2Fl2G73kX0EaAqN5EALw_wcB&gclsrc=aw.ds)
5. [Google Now – ויקיפדיה (wikipedia.org)](https://he.wikipedia.org/wiki/Google_Now)
6. [סירי – ויקיפדיה (wikipedia.org)](https://he.wikipedia.org/wiki/%D7%A1%D7%99%D7%A8%D7%99)
7. [Amazon Lex – AWS Chatbot AI](https://aws.amazon.com/lex/)
8. [אמזון אלכסה – ויקיפדיה (wikipedia.org)](https://he.wikipedia.org/wiki/%D7%90%D7%9E%D7%96%D7%95%D7%9F_%D7%90%D7%9C%D7%9B%D7%A1%D7%94)
9. <https://www.naya-college.co.il/what-is-machine-learning/>
10. [עיבוד שפה טבעית – ויקיפדיה (wikipedia.org)](https://he.wikipedia.org/wiki/%D7%A2%D7%99%D7%91%D7%95%D7%93_%D7%A9%D7%A4%D7%94_%D7%98%D7%91%D7%A2%D7%99%D7%AA)
11. [מה זה למידת מכונה (Machine Learning)? - BDO](https://www.bdo.co.il/he-il/%D7%91%D7%9C%D7%95%D7%92%D7%99%D7%9D/bdo-tech-hub/%D7%9E%D7%A8%D7%A5-2019/%D7%9C%D7%9E%D7%99%D7%93%D7%AA-%D7%9E%D7%9B%D7%95%D7%A0%D7%94-(machine-learning))