

1. מאמר זה בא להציע אלגוריתם (למעשה כותבי המאמר מפנים לאפליקציה שהם כתבו אבל הלב שלה הוא האלגוריתם) ליצירת מוזיקה פוליפונית - צירוף של שני קווים מלודיים עצמאיים (או יותר) תוך יצירה של מרקם הרמוני ביניהם. האלגוריתם מתבסס על generative adversarial networks – GAN. הבעיה מעניינת, ראשית, כיוון שבאופן כללי, יצירה של יצירות ריאליסטיות ואסטרטיות נחשבת לאחד הדברים המרגשים בתחום ה-AI. שנית, בניגוד לתחומים כגון יצירת תמונות, קטעי אודיו וקטעי טקסט בהם יש התקדמות גדולה, בתחום המוזיקה הנושא עדיין לוטה מעט בערפל.
2. פתרונות שהיו קיימים לפני מאמר זה:  
**פתרונות אלו השתמשו ב Recurrent neural network – RNN**
  - Sturm et al 2016
  - Hadjeres, Pachet, and Nielsen 2017 - הצליחו לייצר מוזיקה מונופונית.
  - Boulanger-Lewandowski, Bengio, and Vincent 2012 - פתרון זה הצליח לייצר אמנם מוזיקה פוליפונית אך רק עבור פסנתר בודד.
  - Chu, Urtasun, and Fidler 2017 - פתרון זה הצליח לייצר תפקיד ראשי בודד ולציין את האקורדים המתאימים לליווי התפקיד (קרי - מוזיקה הומופונית).**פתרונות נוספים**
  - Mogren 2016 – שימוש ב RNN+GAN - פתרון זה הצליח ליצר מוזיקה פוליפונית ע"י שילוב של 2 גישות, RNN – בעזרת רשתות כאלה הוא הצליח לייצג מוזיקה כסדרה של אירועים הניתנים לתיאור כרביעיות סדורות של >תחילת הזמן, גובה הצליל, מהירות, משך<. ואח"כ ע"י הצגה של כמה רביעיות כאלה כ- מחוללות (Generator), וכמפלול (Discriminator) במודל GAN. המאמר מציין בעצמו, שאמנם הצליחו לייצר מוזיקה "פוליפונית" ע"ב העקרונות המוזיקאליים, אבל בפועל "הציון" על רמת הפוליפוניה היה נמוך. ברוב המקרים היו מנוגנים תווים בטונים שונים אך למעשה בהפרשים של אוקטבות, כך שלא באמת נוצרה מלודיה שונה ומקבילה למלודיות הנוספות.
  - MidiNet (Yang, Chou, and Yang 2017) - פתרון זה הצליח לייצר מנגינות אבל רק כתלות ברצף אקורדים ידוע מראש, או כהמשך של מנגינה ידועה.
3. המאמר מציע:
  - מודל חדש המבוסס על GAN עבור ריבוי מסלולים (פוליפוניה).
  - יישום המודל המוצע ליצירת מוזיקה, המייצגת, (למיטב ידיעת הכותבים), את הדגם הראשון שיכול ליצור מוזיקה רב-מסלולית, פוליפונית.
  - הרחבה של המודל המוצע כדי לעקוב אחר התניות, קרי, שיתוף פעולה או ליווי מוזיקה עבור מוזיקה נתונה מראש או בשיתוף אנושי.
  - המאמר מציג את Lakh Pianoroll Dataset (LPD), המכיל 173,997 רצועות פסנתר ייחודיות מרובות-מסלולים הנגזרות ממערך Lakh Midi Dataset (Raffel 2016).
  - מספר פרמטרים אובייקטיביים שנוצרו להערכת מוזיקה סימבולית מלאכותית.**תרומת המאמר:**

המאמר מציג מודל גנרי חדשני עבור מוזיקה פוליפונית<sup>1</sup> רצופה במסגרת של GANs. המאמר מעצב מספר ערכים אובייקטיביים (מנגינות) ומראה שניתן לקבל תובנות לגבי תהליך הלמידה באמצעות ערכים אובייקטיביים אלה. המדדים האובייקטיביים ומחקר המשתמש הסובייקטיבי מראים כי המודלים המוצעים יכולים להתחיל ללמוד משהו על מוזיקה. למרות שמבחינה מוסיקלית ואסתטית זה עדיין ירוד ביחס לרמת נגנים אנושיים, למודל המוצע יש כמה תכונות רצויות, שבעזרת מעקב, מחקר ופיתוח יכול להשתפר עוד יותר.
4. כאן אני רוצה מעט להרחיב ולהעמיק בפתרון. המאמר מציע שלוש גישות לפתרון הבעיה, כולן מבוססות GAN<sup>2</sup>. מדובר בשלושה מודלים המתארים תהליכי הלחנה מקובלים, כאשר ההבדל ביניהם הוא בעיקר בתהליך היוצר. התהליך המסווג די אחיד עבור שלושת המודלים.

<sup>1</sup> מוזיקה פוליפונית היא אתגר. תורת ה- קונטרפונקט היא עמוקה וקשה, מלחינים בודדים בהיסטוריה הצליחו להביא את התורה הזאת לידי מיצוי מרבי. כל שכן כשאנחנו באים לדבר על ללמוד מכונה להלחין קטע חדש. ברצוני לדייק כאן נקודה עדינה, קונטרפונקט היא תורה המדברת על היכולת של שתי מנגינות שונות (ויותר) לדור בכפיפה אחת ועדיין להישמע כמשהו הגיוני ואפילו מרגש. למעשה היכולת לשלב מנגינות שונות גובלת ביכולת מתמטית טהורה והדבר המדהים שבה היא היכולת להביא את המתמטיקה הזאת להביע רגש וכוונה. כאשר הכוונה במאמר שלנו היא קצת שונה ממה שדיברנו. המאמר מתעמת עם האתגר בו צריך לקחת מספר כלים שונים, בעלי אופי שונה, גוון ייחודי, תפקידים מסורתיים שונים במסגרת הלחנה, לשלב את כולם במסגרת זמן נכונה וליצור מכולם קטע נגינה אחיד ורציף שגם ישמע "טוב". אז מצד אחד יש כאן עניין של שילוב מרכיבים שונים כדי לייצר מנגינה אחת, אבל עדיין (לפי הבנתי) אין כאן מיצוי של הנושא הפוליפוני.

<sup>2</sup> GAN – אלו למעשה רשתות אדורדיות גטריות. רשתות אלו הן למעשה ארכיטקטורת רשת עצבית עמוקה (CNN) המורכבת משתי רשתות המלמדות/משפיעות אחת על השנייה ולעיתים יקראו "יריבות". הפוטנציאל של GAN הוא עצום כי למעשה ניתן ללמוד רשתות כאלו לחקות כמעט כל הפצה של נתונים. מבחינתנו הפצה של נתונים זו יצירת מוזיקה. הפלט של רשתות אלו הוא בהחלט מרשים. אז איך זה עובד? (ממש בקצרה) GAN מורכב משני אלגוריתמים: אלגוריתם מפלה ואלגוריתם יוצר. המשך בעמוד הבא.

מודל ה **jamming** – דומה להשמה של חברי להקה בחדר אחד וכל אחד מייצר את הלחן הרלוונטי לכלי שלו בשאיפה להשתלב למשהו שיוצר שיר אחיד.

מודל ה **composer** – דומה למלחין בודד המלחין יצירה מקצה לקצה על כל כליה השונים.

המודל ה **היברידי** – מודל שמנסה לשלב את שני המודלים הראשונים.

תהליך העבודה בכל אחד מהמודלים הוא די דומה ולכן נתאר כאן תהליך בדיד ונרחיב אותו למודלים השונים:

יצירת ווקטור רנדומלי  $z$  < האלגוריתם היוצר מקבל את הווקטור ומייצר בעזרתו מנגינה < המנגינה נבחנת ע"י האלגוריתם המפלה. אם היא התקבלה כאמינה היצירה מוכנה, אם לא היא חוזרת שוב לאלגוריתם היוצר לשיפוץ. וכן על זה הדרך. ההבדל העיקרי בין שני המודלים הראשונים הוא בניסיון לדמות את המציאות. כיוון שבמציאות במודל הראשון מדובר על מספר מלחינים, נייצר  $i$  ווקטורים שונים עבור כל מלחין, ולכל אחד יהיה  $i$  generator & discriminator משלו. לעומת זאת, במודל השני כיוון שמדובר על מלחין יחיד, נייצר ווקטור יחיד, generator יחיד, ו-discriminator יחיד. בדיוק כמו שאדם יחיד מקבל החלטה על כל התהליך.

המודל ה **היברידי** שמנסה לשלב בין השניים מייצר ווקטור "ראשי" עבור המלחין הראשי, הווקטור הזה נותן "היסט" (במשמעות של השפעה) לווקטורים המשניים של שאר הנגנים (נראה לי שע"י חיבור של כל ווקטור עם הווקטור הראשי וכך יוצא שכולם "מתקרבים" אל הווקטור של המלחין הראשי), כולם משתתפים בהלחנה ולכן ישנם מספר generator ים- כמספר המלחינים (ללא המלחין הראשי) ובסוף discriminator אחד כי מי שמקבל את ההחלטות בסוף הוא המלחין הראשי.

### הדרך בה הם בדקו את תוצאות ההלחנה (יצירת הפרמטרים האובייקטיביים):

כותבי המאמר קבעו מדדים אובייקטיביים לפיהם הם יבחנו את המנגינות. המדדים הם:

- יחס התיבות הריקות. (%)
- מספר הפעילות בהן משתמשים בכל תיבה. (0-12)
- אורך תו מנוגן (הוא צריך להיות לפחות באורך מסוים. מדד זה בא להגיד בשפה חופשי כמה המנגינה "מקוטעת") (%).
- בדיקת יחס פעילות לתיבה (%) (לדוגמה בשירי רוק מקובל לרוב היחס 4/4)
- בדיקת מרחק טונלי בין המסלולים. כאשר מרחק גדול יותר מייצג קשר טונלי הרמוני חלש יותר.
- אחרי שנקבעו המדדים, ה-discriminator עובר על כל ה-dataset לפי המדדים הללו ומכוון את עצמו לזהות מהו לחן "טוב". כעת הוא ידע להגדיר האם ה-generator הלחין משהו תקין או שצריך להלחין מחדש. בסופו של דבר, אחרי שכיוונו את כל המשקולות והצלעות של ה-generator ע"י ה-discriminator, ביחס ל-dataset, ניתן ל-generator ליצור מנגינות במודלים השונים ונמדדו לפי המדדים שנקבעו. כך נוכל לראות האם האלגוריתם אכן הצליח ללמוד משהו ולהתקרב למדדים שנקבעו ומתאימים ליצירה אמיתית.

### 5. עבודה לעתיד:

- לנסות לחצות את גבול 4 התיבות ולהלחין שיר מלא (אפרט בכיתה מה האתגר בכך).
- Cross-Model Generation:
- יצירה של מוזיקה ואודיו
- יצירה של מוזיקה עם טקסט מתאים

### 6. סיכום:

אני נהניתי מאוד מהמאמר, אמנם הרבה בגלל ההיבטים שקשורים אלי כמוזיקאי והיכולת לשלב את המקצוע עם התחביב, אבל גם מלמידת הרעיון של GAN כרעיון כללי ואת המימוש שלו כאן באופן פרטי. המאמר פתח בפני הרבה סוגיות לחשוב עליהן, כגון מה הסיבה בכלל לרצות שמחשב ילחין משהו? איך זה יכול לעזור למישהו? עד לאן אנחנו רוצים להביא את יכולות המחשב? האם מוזיקה היא לא נחלת בני אנוש בלבד? איך ניתן לקחת משהו בעל מטען רגשי אדיר ולהפוך אותו למכונה? הרבה מחשבות אפילו מעבר ללמידה שקשורה למדעי המחשב.

---

תפקיד **האלגוריתם המפלה** הוא לענות על השאלה הבאה, בהינתן שאני יודע את  $x$  מה ההסתברות ש  $y$  ( $P(y|x)$ ). לדוגמה נניח שיש לי סט של מילים שקיבלתי במייל, מה ההסתברות שהמייל הזה הוא spam? למעשה האלגוריתם המפלה יודע לתת תווית למאורע. הוא מסתכל על התכונות שלימדו אותו להסתכל עליהן באימון (סט המילים המרכיבים מייל spam), ולהכריע האם המייל הנתון ראוי לשאת את התווית spam. ניתן לחשוב על תפקיד **האלגוריתם היוצר כתפקיד ההופכי לאלגוריתם המפלה**. דהיינו בהינתן שאני יודע שמייל מסוים הוא spam, האם אני יודע לייצר סט של מילים המרכיבות מייל spam. שזה כמו לענות על השאלה  $P(x|y)$ . שני האלגוריתמים עובדים בצורה סימולטנית כאשר ה"אלגוריתם היוצר" יוצר מופעים חדשים ואילו האלגוריתם המפלה מסווג אותם ע"פ מערך האימון כמופעים אותנטיים או לא. נניח שאנחנו מנסים לעשות משהו בבנאלי יותר מאשר לחקות מוזיקה פוליפונית. אנחנו רוצים לייצר ספרות בכתב יד כמו אלה שנמצאו במערך הנתונים MNIST, אשר נלקח מן העולם האמיתי. מטרתו של המפלה, כאשר מוצג דוגמה ממערכת הנתונים האמיתית של MNIST, היא לזהות אותם כאותנטיים. בינתיים, הגנרטור יוצר דימויים חדשים שהוא מעביר למפלה. היא עושה זאת בתקווה שגם הם ייחשבו אותנטיים, למרות שהם מזויפים. מטרתו של הגנרטור היא ליצור ספרות כתובות ידיניות, לשקר בלי להיתפס. מטרתו של המפלה היא לזהות דימויים שמקורם בגנרטור כזיוף. שתי הרשתות מנסות לייעל פונקציה אובייקטיבית אחרת ומנוגדת, או פונקציית הפסד, במשחק סכום-אפס. זהו למעשה מודל של actor-critic. כאשר המפלה משנה את התנהגותו, כך גם הגנרטור, ולהיפך. למעשה ההפסדים שלהם דוחפים אחד את השני.