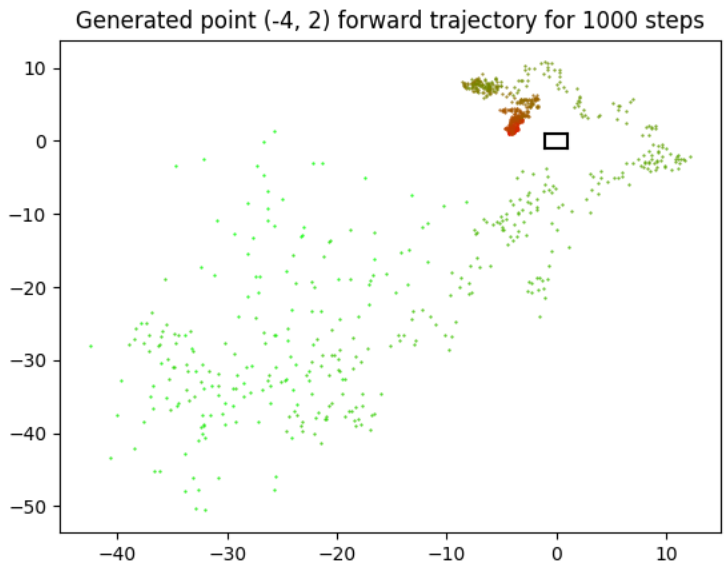
תרגיל 1 – מודלים גנרטיביים

67912 - קורס מתקדם בלמידת מכונה

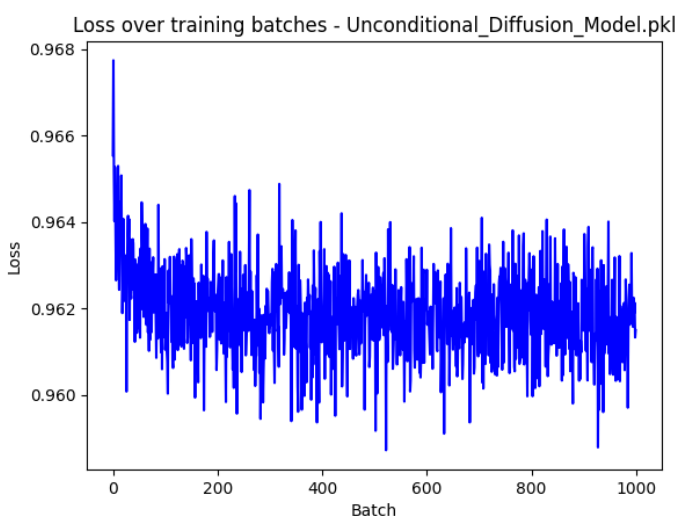
אסף שול 207042714

חלק א' – Diffusion Models

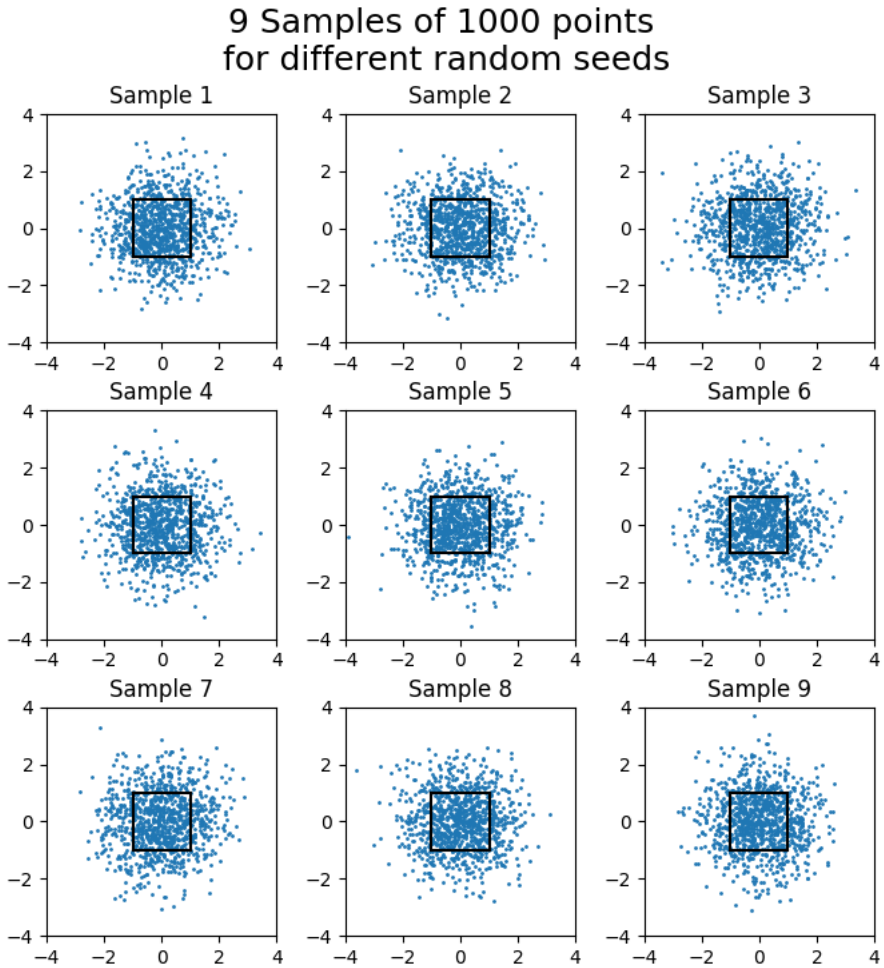
1. גרף שמתאר את תהליך ההרעשה של נקודה אקראית כתלות בt, מ0 עד 1, כאשר הרעש נוסף בכל שלב לרעש הקודם בצורה איטרטיבית:



2. גרף השגיאה באימון המודל: להחליף לעדכנייייי

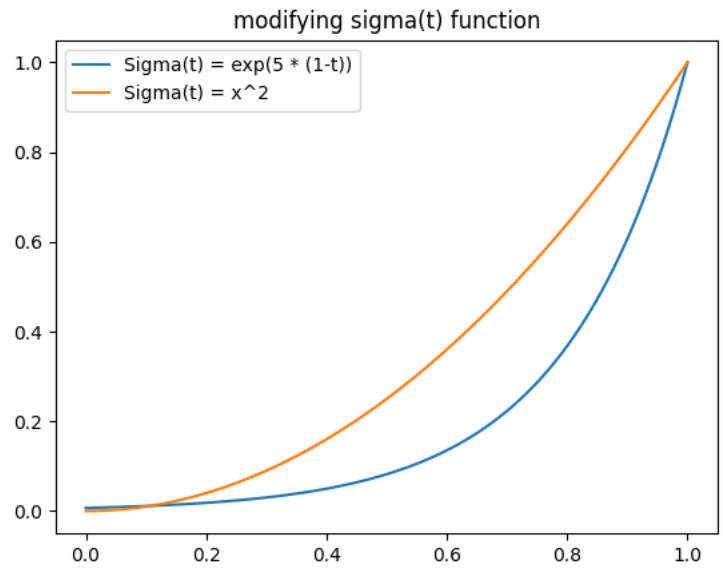


3. גרף 9 דגימות של 1000 נקודות:

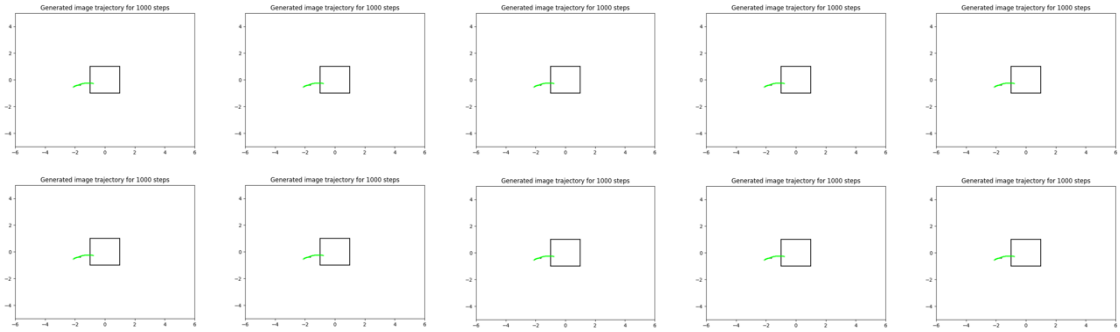


4. השפעת t על הדגימה?

5. שינוי הסמפלר?

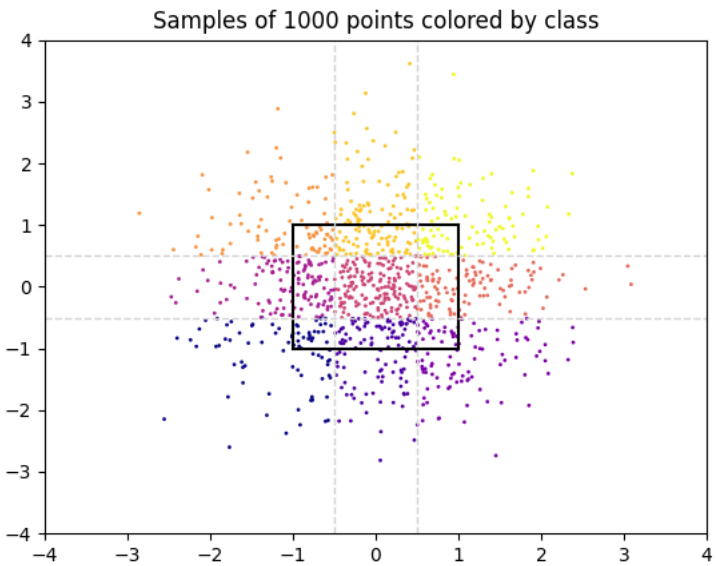


6. ניתן לראות שבהרצת המודל 10 פעמים, עם אותו רעש, נקבל בדיוק את אותה תוצאה:

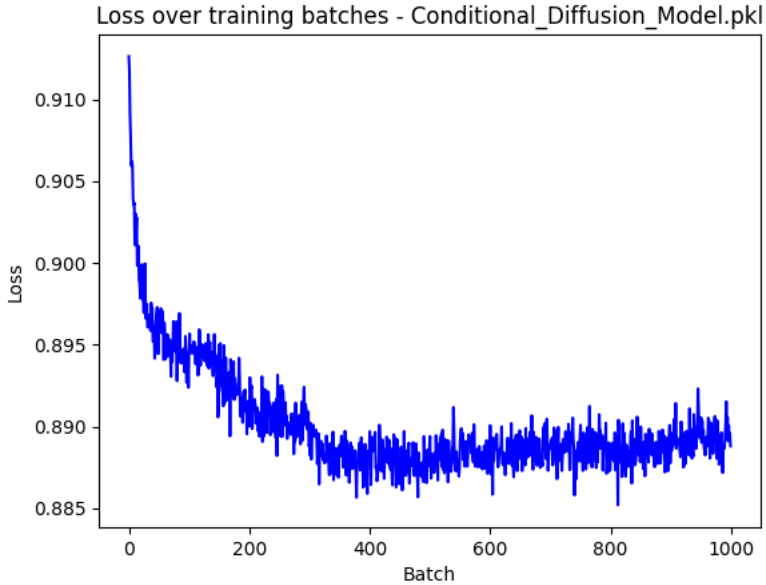


...שינוי במודל

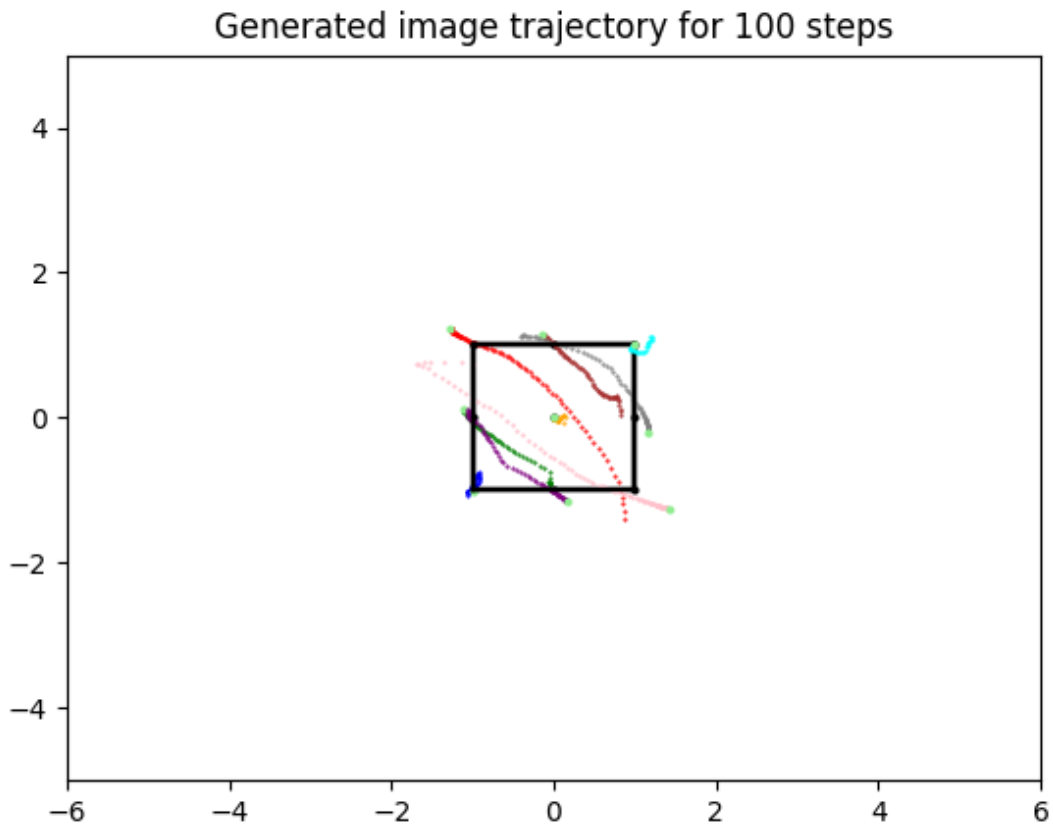
.1


Samples of 1000 points colored by class
4
3
2
1
0
−1
−2
−3
−4
−4
−3
−2
−1
0
1
2
3
4

.2


Loss over training batches - Conditional_Diffusion_Model.pkl
Loss
0.910
0.905
0.900
0.895
0.890
0.885
0
200
400
600
800
1000
Batch

.4


Generated image trajectory for 100 steps
4
2
0
−2
−4
−6
−4
−2
0
2
4
6

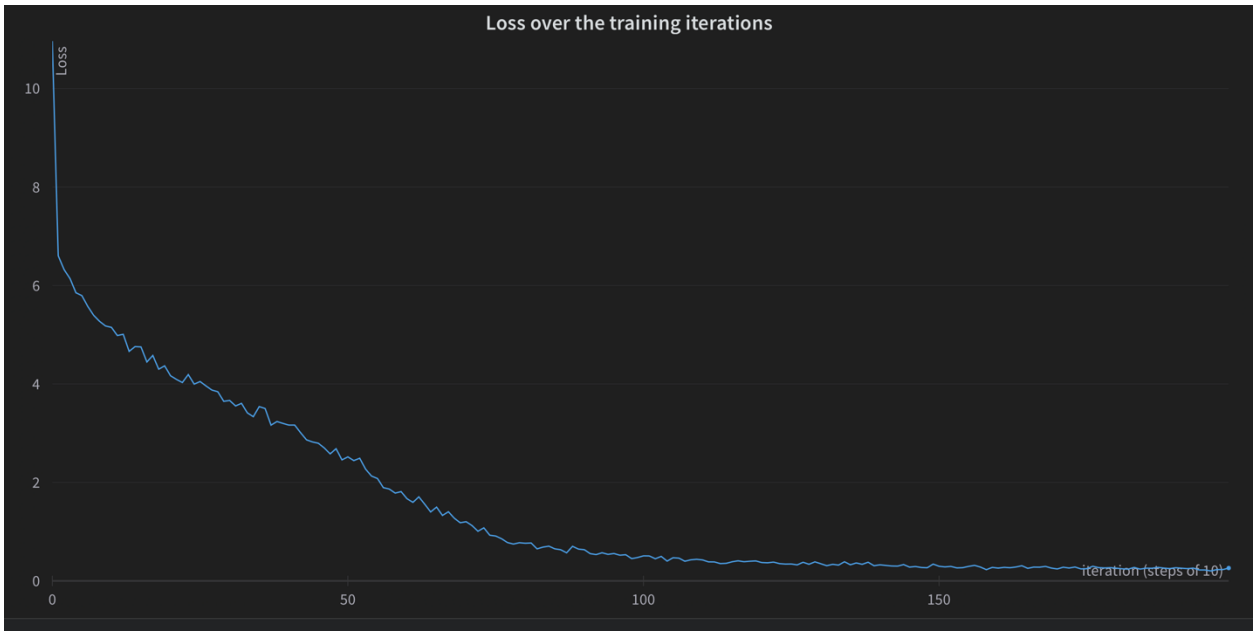
חלק ב' – Auto-Regressive Text Models

1. אימון המודל:

המודל אומן על הטקסט של הספר "אליס בארץ הפלאות" עם הפרמטרים הבאים:

- Block size: 64
- Batch size: 32
- Learning rate: 5e-4
- Iterations number: 2000

ניתן לראות את הלוס של המודל לאורך הזמן בגרף הבא:



2. כפי שהשאלה מורה – ניסיתי לגרום למודל לפלוט את המשפט:

"I am a little squirrel holding a walnut"

נקודה מעניינת ראשונה ששמתי לב אליה בעת ביצוע משימה זו, היא שבקורפוס כולו אין את המילים "סנאי" ו-"אגוז" כלל.

מכיוון שלכל מילה הסתברות חיובית ממש בוקטור הפלט של המודל, אם נורה למודל לדגום ע"פ וקטור ההסתברות ולא פשוט לקחת את המילה בעלת ההסתברות המקסימלית, על הנייר ניתן לקבל משפט זה. דרך זו התבררה מהר מאוד כלא פיזבילית שכן הסיכוי לדגום את אחת המילים, ובטח שאת המשפט כולו, הוא לא משהו בר ביצוע.

לאחר מכן ניסיתי להכניס למודל קלט של תופעה שקיוויתי שהוא ידע להכליל – פרומפט בסגנון של:

```
Repeat after me: 'I am a little  squirrel holding a walnut'. And Alice
said:
```

אבלך גם דרך זו לא צלחה, ניסיתי לראות האם קיים בטקסט פורמט שאכן אומרים לאליס לחזור אחר משפט מסוים והיא חוזרת אחריו בצורה זהה, ואכן מצאתי כמה מקרים כאלו, ערכתי אותם לצרכי וניסיתי לראות את פלט המודל עליהן:

```
Repeat, "I AM A LITTLE, SQUIRREL HOLDING, A WALNUT,"' said the
Caterpillar. Alice folded her hands, and began:--
```

או

```
The Fish-Footman began by producing from under his arm a great
letter, nearly as large as himself, and this he handed over to
the other, saying, in a solemn tone, `I am a little squirrel holding a
walnut.' The Frog-Footman repeated, in the same solemn tone, only changing
the order of the words a little
```

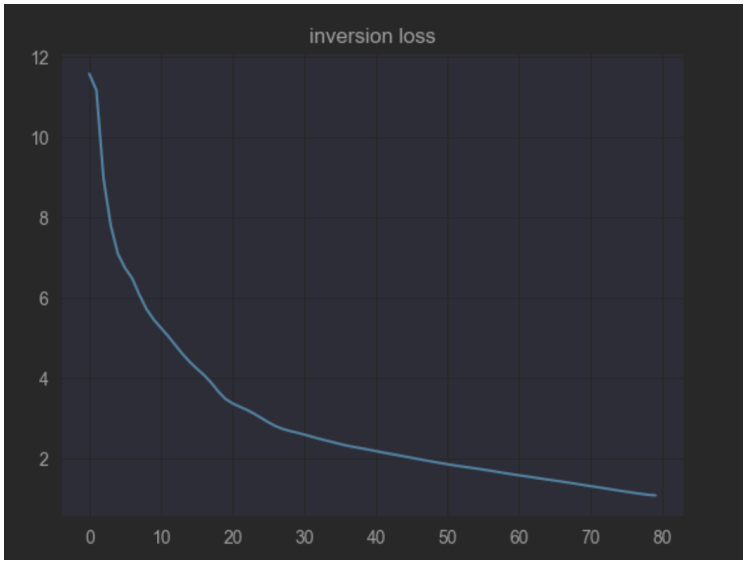
אך גם ניסיונות אלו כשלו, כשאימנתי את המודל למשך זמן רב יותר (יותר איטרציות ובאטצ'ים גדולים יותר) הוא פשוט שינן את המשפטים המקוריים מהדפר וענה כמוהם, ואילו כשאימנתי פחות – לא הצליח להכליל את תופעת ה-"חזור אחרי" שקיוויתי שיכליל.

לאחר מכן ניסיתי לפתור את הב

לאחר מכן ניסיתי לפתור את הבעיה בדומה לדרך בא ניסינו למצוא את הווקטור הלטנטי המתאים לתמונה ספציפית בתרגיל 5 בקורס עיבוד תמונה, וביצעתי gradient decent עם מודל מאומן, כשהתחלתי מווקטור אקראי. בשיטה זו סוף סוף הצלחתי לגרום למודל לפלוט את המשפט הנכון, אצרף תמונה של התהליך האטרקטיבי, גרף הלוס והקלט שהתקבל:

```
  0%|          | 0/1000 [00:00<?, ?it/s]
current sentence for latent:  now out on his now of" for so
  3%|          | 26/1000 [00:04<02:32,  6.38it/s]
current sentence for latent: quet am a littleney holding a wal?
  5%|          | 51/1000 [00:08<02:29,  6.35it/s]
current sentence for latent: I am a little birds holding a walnut
  8%|          | 76/1000 [00:12<02:27,  6.26it/s]
current sentence for latent: I am a little clever holding a walnut
  8%|          | 79/1000 [00:13<02:33,  6.01it/s]
--------------------------------------------------
Stopped !
model output: " I am a little squirrel holding a walnut "
```

```
tensor([[[ 0.8371, -2.7097,  2.5457,  ..., -2.5255,  0.9687, -1.0763],
         [-0.7460,  1.4124,  1.6422,  ..., -1.0766,  0.3521,  0.2970],
         [-0.6000, -0.1209, -0.4295,  ...,  1.5853, -0.6543,  1.8726],
         ...,
         [ 0.8655, -1.3813,  0.2538,  ..., -1.2783,  1.2740,  0.2395],
         [-1.7015, -1.1999, -0.3732,  ...,  0.8111,  1.1715, -1.4626],
         [-2.9633, -0.1542, -3.4524,  ...,  0.8937,  0.2437,  0.4835]]],
       requires_grad=True)
```



3. ו 4.

בשאלות שלוש וארבע נתבקשנו לחקור את אופן ההתנהגות וההשפעה של הattention במודל שלנו, אענה עליהן יחד בכדי להציג באופן טוב יותר את נקודות הדמיון והשוני בין הattention של הבלוק הראשון ושל הבלוק האחרון.

לצורך בדיקה זו יצרתי משפט אקראי בן 10 מילים:

"The little squirrel was around Alice but she kept saying"

(כפי שניתן לראות מוטיב הסנאים מהשאלה הקודמת ממשיך)

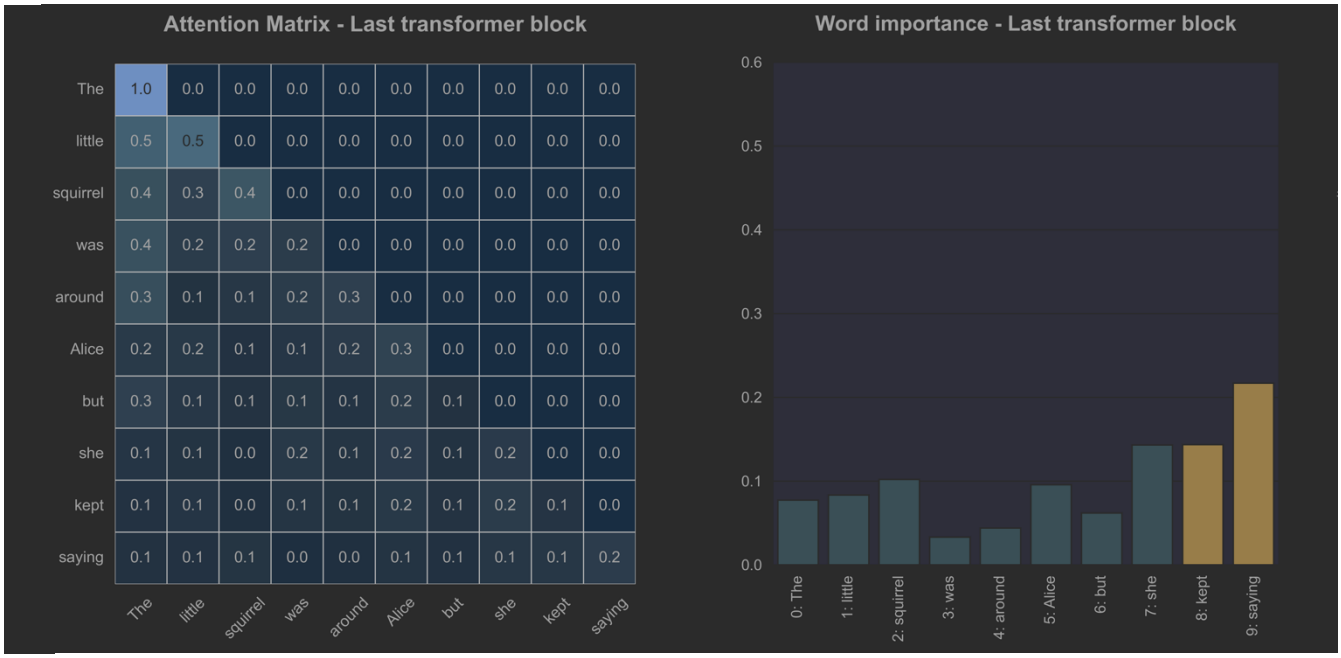
כעת הרצתי את המודל עם משפט זה כקלט, ושמרתי את מטריצות הattention שלו בעת ביצוע הפרדיקציה למילה ה11. את המטריצה שמרתי כממוצע של כל 12 הראשים של ה-multi-attention במודל. בגרפים הבאים ניתן לראות את מטריצות ה-attention של בלוק הטרנספורמר הראשון והאחרון, ואת החשיבות שנתן לכל מילה בעת ביצוע הפרדיקציה (2 הכניסות עם הערכים הגבוהים ביותר מודגשים בצבע כתום) *מצורפים בעמוד הבא.

ניתן לראות שבבלוק הראשון, המודל נתן את רוב המשקל בattention למילה האחרונה בקלט (אלכסון מטריצת הattention) ואילו בבלוק האחרון הפיזור ניראה יותר שוויוני.

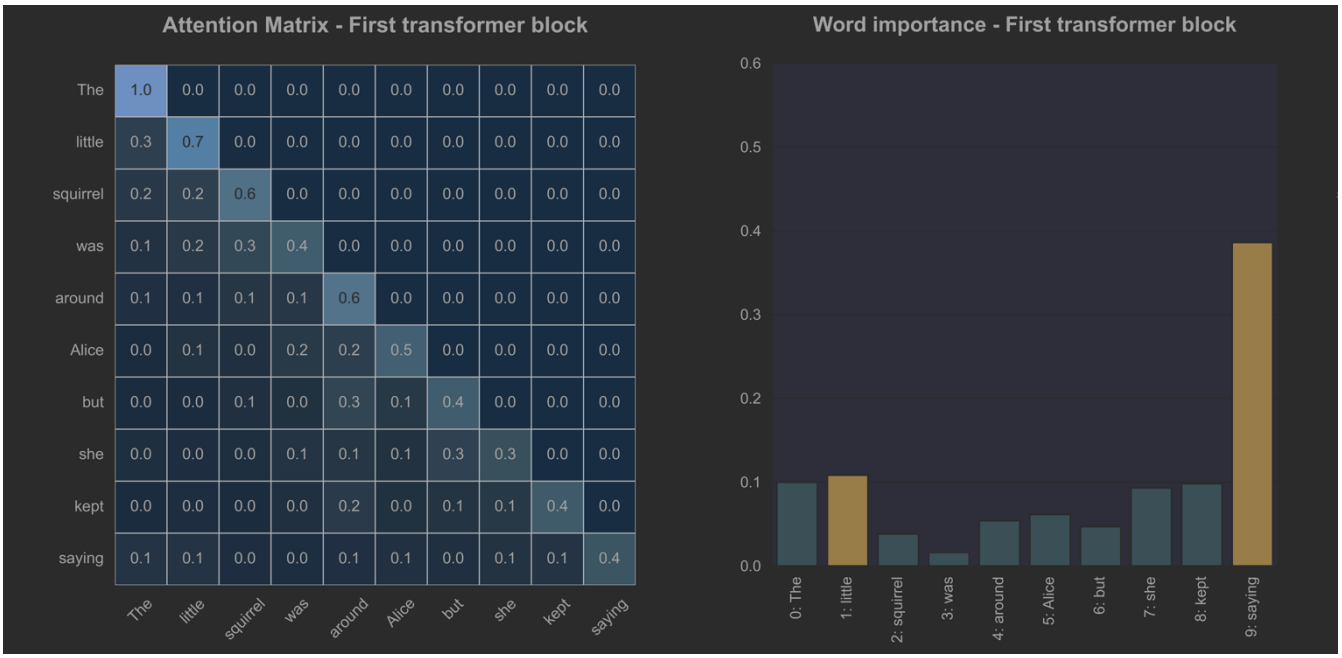
אני מסיק מכך ככל שאנחנו נמצאים בבלוק עמוק יותר במודל, כל כניסה היא כבר קומבינציה לא לינארית של הכניסות הקודמות, עם משקולות מסוימים, ולדעתי הדבר מביא לייצוג ברמה סמנטית גבוהה יותר, ולכן המידע החשוב למודע ברמה הסמנטית נמצא מפולג באופן אחיד יותר לאורך הכניסות.

כשהתייעצתי על כך עם גבי בשיעור נל"פ מתקדם, הוא אמר שלפעמים ניתן לראות תופעה במודלי שפה שדומה לתופעות שראינו במודלי תמונה, שהבלוקים הראשונית נוטים לתפוס ייצוגים של דברים פשוטים יותר כמו חלקי הדיבר לדוגמא (שקול רעיונית לקרנלים הראשונים ברשת קונבולוציה שלומדים צורות פשוטות כמו קווים או פינות) ואילו הבלוקים האחרונים תופסים רעיונות סמנטיים מורכבים יותר.

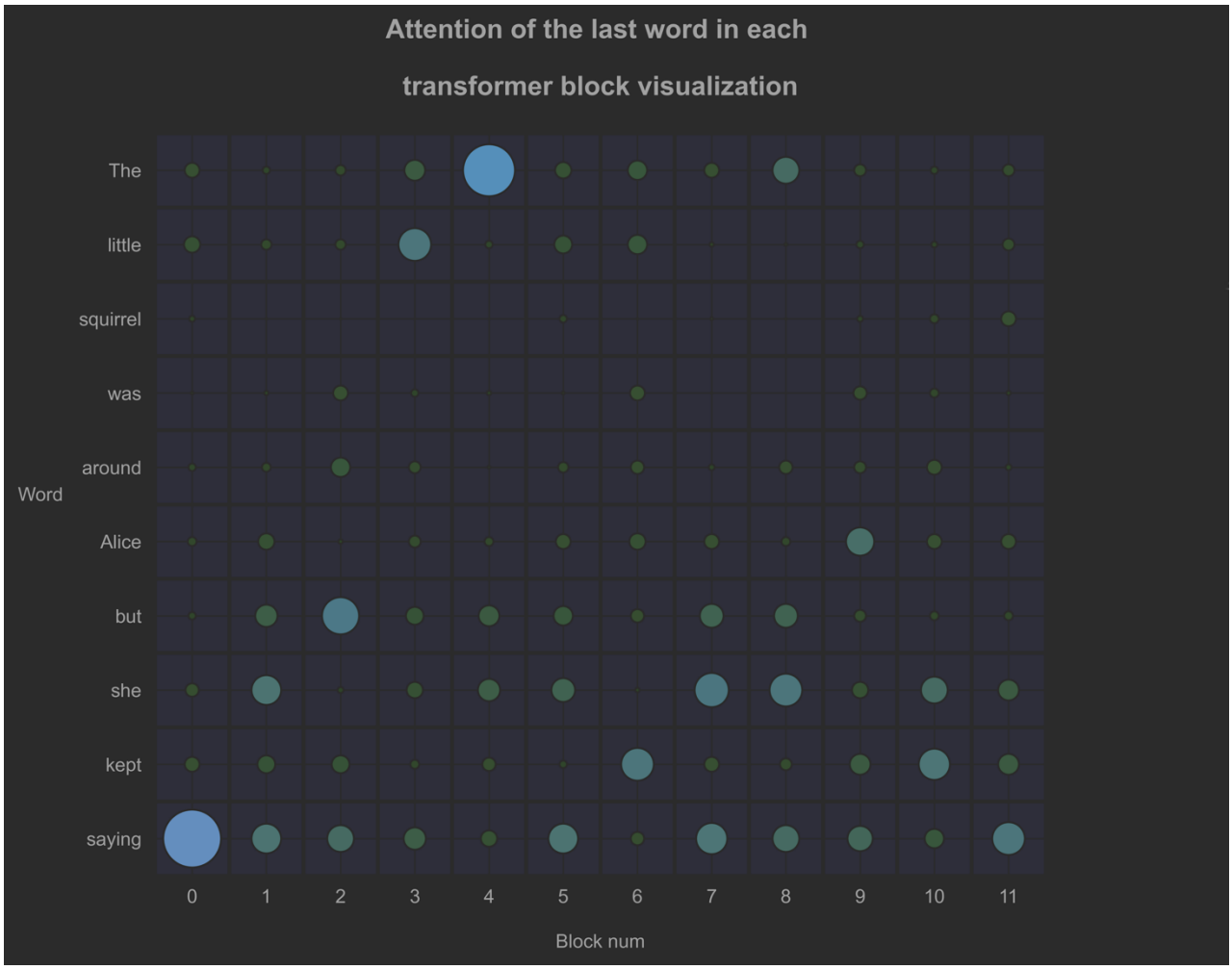
עבור הבלוק האחרון:



ועבור הבלוק הראשון:



בנוסף ניסיתי לייצר ויזואליזציה של חשיבות הכניסה בכל בלוק עבור הפרדיקציה של המילה ה-11 בצורה גרפית נוחה, אצרף זאת כאן. גודל כל עיגול מייצג את החשיבות של הכניסה בבלוק המתאים:



5. ההסתברות לפי המודל למשפט מהשאלה הקודמת, עבור 15 המילים הבאות (מוצג לא לוגריתמי) למשפט שהמודל פלט הייתה הסתברות של כ-0.00042~ והוא היה:

```
The little squirrel was around Alice but she kept saying, `That's
right, Five! I wonder if I shall have got
```

ואילו ההסתברויות היו: