# Review 175: [Short] A Picture is Worth a Thousand Words: Principled Recaptioning Improves Image Generation

**Paper: https://arxiv.org/abs/2310.16656v1**

https://huggingface.co/papers/2310.16656

אוקיי, מכיוון שהסקירה הזו היא סקירה של שבת, אז היא (הסקירה) הולכת להיות יותר קלילה ויותר קצרה מהרגיל. אחרי שאתמול סקרנו מאמר על מודלי שפה היום חוזרים למודלי דיפוזיה. המאמר המסוקר מציע שיטה די פשוטה לשיפור של מודל דיפוזיה טקסטואלי ההופך תיאור טקסטואלי לתמונה. המודל המשופר מצליח לייצר תמונות מתאימות יותר לתיאור הטקסטואלי בצורה מדויקת יותר. השיטה המוצעת מכילה 3 שלבים:

גנרוט של 10K תמונות איכותיות ממודל דיפוזיה רגיל (aesthetics score ≥ 5.0, p\_watermark < 0.5 etc). מודל דיפוזיה רגיל נבחר כמובן Stable Diffusion.

כיול של מודל היוצר כותרת לתמונה (PaLi). כדי לכייל מודל captioning המחברים ביקשו ממתייגים אנושיים לתייג 100 תמונות ולתת לכל אחת 2 כותרות: אחת קצרה ותמציתית והשנייה ארוכה יותר אך עדיין מדויקת. לאחר מכן מודל ה-captioning טיוב עם הדאטהסט הזה

מפעילים את PaLi על הדאטהסט מהשלב הראשון ומכיילים SD על הדאטה הזה.

זהו זה – כך מקבלים מודל דיפוזיה משופר. הבטחתי לכם קל וקצר וקיימתי.