# Review 163: [Short] Idea2Img: Iterative Self-Refinement with GPT-4V(ision) for Automatic Image Design and Generation

**Paper: https://arxiv.org/abs/2310.08541v2**

https://huggingface.co/papers/2310.08541

אוקיי, חברים אחרי הפסקה האורך של יותר מחודש אני חוזר לסקור קצרות של #shorthebrewpapereviews. האמת שעשיתי כמה נסיונות לחזור קודם לכן אבל המוח כנראה לא היה מוכן לכך. מתחילים מהמאמר שיצא שבוע אחרי המלחמה ולאט לאט נתקדם עם המאמרים היותר טריים.

אז המאמר (די קליל כי מוחי לא חזר לעצמו במלוא החדות) שנסקור היום פיתח שיטה המאפשרת לשכלל את היכולות של GPT4-V (שיודע לעבוד עם דאטה ויזואלי) ביצירה תמונה מתיאור רעיוני. כלומר עבור תמונה של אדם ותמונה של כלב הוא מאפשר ליצור תמונה של אותו האדם עם הכלב מהתמונה השנייה. השיטה די אינטואיטיבית ומנצלת יפה את היכולות של מודלי שפה-תמונות משולבים. השיטה מורכבת מ 4 שלבים:

המודל יוצר N פרומפטים מרעיון (IDEA) נתון. ליצירת פרומפטים הרעיון (שיכול להיות תיאור או כמה תמונות) מוזן ל-LMM שזה Language Multimodal Model (לא לבלבל עם LLM).יי כמובן של יש פרומפט קבוע (מטה-פרומפט) שמוזן ל-LMM כדי להסביר לו מה לעשו. לאחר מכן יוצרים N תמונות מהפרומפטים האלו (אפשרי עם מודל אחר). גם ה״רעיון״ מוזן למודל כמובן.

בוחרים את התמונה הכי טובה לפי מידת התאמה לרעיון באמצעות שאילתה ל-LMM (הם אמרו שרק מודלים מתקדמים יודעים לעשות זאת ביעילות רבה).

מזינים את התמונה הנבחרת ל-LMM ומבקשים ממנו לספק לנו משוב על איכותה (של התמונה). בשביל לשפר את איכות משוב שומרים בזיכרון את כל התמונות שנבחרו בשלב 2 כולל הפרומפט שהן נוצרו ממנו. זה משרת כסוג של "גרדיאנט״ בתהליך הזה.

חוזרים על שלבים 1-3 T איטרציות ומקבלים את התמונה מה״רעיון״.

התוצאות די נחמדות