# Review 104: [Short] Evaluating machine comprehension of sketch meaning at different levels of abstraction, 23.07.23 https://cogtoolslab.github.io/pdf/mukherjee\_cogsci\_2023.pdf

**Paper: https://arxiv.org/abs/2312.03035v1**

האם מודלים של ויז'ן ״מבינים״ את עולם הויזואלי בדומה לנו, בני האדם? האם הם ״מבינים״ קונספטים אבסטרקטיים כמו סקיצה של תמונה? היום ו ב-shorthebrewpapereviews# מאמר מגניב, חמוד ולא רגיל הבודק את היכולת של מודלי ויז'ן חזקים להבחין אובייקטים (חיות) על סמך הסקיצות שלהם בלבד.   
  
הם בדקו את זה על סקיצות מפורטות יותר (שמכילים יותר פרטים על אובייקט) וגם כאלו מאוד סכמתיים המכילים כמה קווים שניתן לצייר אותם ב-8 שניות (לא אדם כמוני שלא יודע לצייר אלא מישהו עם קצת יכולות). אנחנו, בני האדם, ברוב המקרים לא מתקשים לזהות אריה גם לפי סקיצה מאוד בסיסית שלו.   
  
לעומת הזאת המודלים החזקים שלנו כמו CLIP או VIT-B די מתבלבלים עם הסקיצות ומתקשים לזהות אובייקטים בסקיצות שלהם. כמובן אתם שהמודלים האלו לא אומנו על הסקיצות ולא צפוים לעבוד טוב עליהם, עדיין ההבנה הזו היא נחמדה.   
  
המחברים גם בדקו האם הסקיצות שנוצרות באמצעות מודל גנרטיבי(CLIPasso) הם ״מובנים״ על ידי בני האדם כמו הסקיצות הטבעיות. מתברר שהסקיצות המצוירות על ידי מובנות לבני אדם באותה רמה כמו הסקיצות הטבעיות המפורטות (של 30 שניות). לדעתי זו תגלית די מעניינת.