# Review 197: [Short] LLM4Decompile: Decompiling Binary Code with Large Language Models

**Paper: https://arxiv.org/abs/2403.05286v3**

https://arxiv.org/abs/2403.05286

המאמר מציע LLM4Decompile, משפחה של מודלי LLM לדיקומפילציה בגישה פתוחה שנעים מ-1B עד 33B פרמטרים. מודלים אלו מאומנים על 4 מיליארד טוקנים של קוד מקור בשפת C וקוד אסמבלי מתאים. המחברים גם מציגים את Decompile-Eval, הדאטהסט להערכת דיוק הדקומפילציה המבוצעת על ידי מודל (מקמפלים מחדש ובודקים את הפונקציונליות של הקוד).

LLM4Decompile מצליח לבצע דקומפילציה בצורה מדויקת 21% מקוד האסמבלי, עם שיפור של 50% ביחס ל-GPT-4. מודלי שפה לקימפול ולדקמפול של קוד נראים כמו תחום מחקר חשוב במיוחד לבניית סוכני AI חסינים יותר נגד התקפות אדוורסריות שיפעלו בשכבות העמוקות יותר של Software Stack.