**תקציר עבודת גמר לתחרות מדענים צעירים**

שם העבודה: דגימה בקבוצות – שימוש יעיל במודלי שפה סיבתיים (causal language models)

שם מלא: יוני קרמר

בית ספר: תיכון עירוני ד' על שם אהרון קציר, תל אביב

מנחה ומורה מלווה: מר עידו גודיס

מנחים מטעם התחרות: גברת ליטל שיריין ומר צביאל למברגר.

מודלי שפה סיבתיים כגון chat GPT מבצעים משימות בינה מלאכותית רבות בהן נקלט ונפלט טקסט כגון: שיחה, תרגום, סיכום, מענה על שאלות, ותכנות.

באלגוריתם האוטורגרסיבי, המודלים מוצאים את המילה הבאה בהינתן רצף מילים. המודל מקבל את הקלט ואת המילים שמצא עד כה ומוצא את המילה הבאה בפלט כך שיצירת n מילים מחייבת n שימושים במודל מה שגורם לחוסר יעילות ולעלות מחשוב גבוהה מאוד.

בעבודה פותח אלגוריתם בשם דגימה בקבוצות היוצר טקסטים ארוכים בשימוש אחד בלבד במודל שפה סיבתי קיים. הניסוים בעבודה הראו שדגימה בקבוצות מצליחה בתרגום יותר מהאלגוריתם הקיים ושהיא זולה ויעילה פי 1182 מהאלגוריתם הקיים.