**השפה של ציטוטים זכורים (מאמר).**

הדאטה – ציטוטים זכורים (לפי הופעות אינטרנט + הופעה ב-IMDB) מ-1000 סרטים (קורפוס הסרטים הגדול). בנוסף גם ציטוטים לא זכורים מאותה סביבה.

אנשים אמרו שמה שהופך ציטוטים לזכירים זה כלליות שלהם, וייחודיות שלהם.

מה שעשו במאמר:

לקחו language model כלשהו, והראו שאכן הציטוטים הזכירים היו בעלי לייקליהוד נמוך יותר לפי המודל, מאשר אלו הלא זכירים (כלומר, הם אכן ייחודיים).

**מה שהם עשו זה לזרוק את הציטוטים שיש להם יותר משורה אחת.**