# Review 94: [Short] In-context Autoencoder for Context Compression in a Large Language Model

**Paper: https://arxiv.org/abs/2307.06945v4**

#llm שלכם לא מבין את טקסטים ארוכים כי אורך הקשרו קצר מדי?

רבים ניסו לפתור : Hyena, RMT, LongNet? אז הנה עוד מאמר אחד שמנסה להשתמש בפתרון הדי מתבקש קרי AutoEncoder

היום ב-#shorthebrewpapereviews:

In-context Autoencoder for Context Compression in a Large Language Model

אז בואו נדחס את הקלט ל-llm בצורה כזו שהוא כן ייכנס לחלון ההקשר של llm. אבל איך לעשות זאת בלי לאבד את התכונות הייחודיות של הקלט? נכון אנו ננצל את הגישה החביבה של AE. אנו נדחס את הדאטה כך עם ה-encoder שה-decoder שלו יידע לפענח את הייצוג הדחוס הכי קרוב למקור.

ואכן כך הם עשו. בשלב הראשון אימנו AE לדחוס את הטקסט. בשלב השני לקחו את AE המאומן כיילו llm להשלים טקסטים. בשלב האחרון כיילו מודל שפה לעקוב לאחר הוראות (instruction fine-tuning). וככה קיבלו llm שיודע לאכול טקסטים ארוכים.

https://arxiv.org/abs/2307.06945