# Review 179: [Short] UNLEASHING THE POWER OF PRE-TRAINED LANGUAGE MODELS FOR OFFLINE REINFORCEMENT LEARNING

**Paper: https://arxiv.org/abs/2310.20587v5**

https://huggingface.co/papers/2310.20587

המאמר הזה משך את עיני כי אני מאוד אוהב שילובים של מודלי שפה (וגם מודלי דיפוזיה) למשימות מהעולם של למידה עם חיזוקים (Reinforcement Learning). המאמר הזה עשה את זה בצורה מאוד אלגנטית כל כך אחרי שרפרפתי בו 5 דקות לא הבנתי חשבתי שזה די בלתי אפשרי. אבל אחרי הקצת צללתי לעומק הבנתי שכדאי לתת לזה צ'אנס ולסקור אותו במדורנו.

אז מה המאמר עשה בעצם? בגדול הם לקחו מודל שפה וטייבו (finetuned) אותו לבצע למשימות של RL. כלומר בהינתן של פעולות ומצבים קודמים המטרה של המודל היא לחזות את הפעולה הבאה. במקרה הזה מדובר באופליין RL כלומר המטרה של המודל היא לחקות כמה שיותר טוב את הפעולות המוצלחות מהדאטהסט (בהינתן הפעולות והמצבים הקודמים). במשימות שנדונו במאמר הפעולות מתוארות בצורה מילולית.

כבר מריחים את מודלי השפה מתקרבים? אוקיי, קודם המחברים לקחו מודל שפה מאומן (GPT2) וטייבו אותו על הדאטהסט הנקרא WiKiText. בשלב השני מוסיפים למודל שפה כמה שכבות של MLP ומאמנים אותו על הדאטה של המשימה (נגיד משחק אטארי) ובנוסף מכיילים מודל שפה עם LoRA (זוכרים מה זה?). תוך כדי התהליך הזה מוסיפים איבר רגולריזציה המכיל לוס על הדאטה של WikiText כנראה כדי לגרום למודל לא לשכוח את המיומנויות הקודמות שלו. וזה וזה מקבלים מודל ל-RL כלומר decision transformer עם ביצועים טובים.