# Review 115: [Short] From Sparse to Soft Mixtures of Experts, 03.08.23 https://huggingface.co/papers/2308.00951

**Paper: https://arxiv.org/abs/2308.00951v2**

ערבוב של מומחים (mixture of experts) – שמעתם על זה? כשיש לכם משימה מורכבת ביד אחת הגישות לפתור אותם היא לחלק אותה לכמה תת-משימות בעלות אופי שונה ואז מאמנים מודל לכל אחד מהמשימות אלו. בסוף משלבים את התוצאות של כל מודלים לבניית הפתרון לבעיה המורכבת שלנו.

המאמר שנסקור היום ב-#shorthebrewpapereviews סוקרים מאמר המציע גישה חדשה, פשוטה ואלגנטית לביצוע MoE המאפשרת להקטין את כמות החישוב הנדרש לאימון המודל למודלים עצומים בגודל (בגדול טרנספורמרים). בעבר שיטות MoE k דלילות (sparse) חילקו פיסת דאטה (נגיד טקסט או תמונה) לכמה חלקים שונים ואז כל מודל מתאמן רק על חלק מפיסת הדאטה וככה היה נחסכת כמות נכבדת של חישובים.

לגישה זו יש כמה חסרונות כמו התפלגות מאוד לא שוויונית של הדאטה בין המודלים המומחים, אימון לא יציב וקושי לעשות סקייל מספר המומחים. הגישה המוצעות מציעה פתרון מאוד אלגנטי לסוגיות אלה – במקום לחלק את פיסת הדאטה בין המודלים המומחים כל אחד מקבל קומבינציה לינארית שונה (או כמה) של כל חלקי הדאטה. כך כל מודל רואה את כל פיסת הדאטה במלואו אבל עם זאת זה מאפשר להקטין משמעותית את מספר החישובים. אהבתי!