***Investigating the Effect of Training Data Order on Small Language Model Fact Retention* - or "*ordering is all you need*"**

- Yair Ben Shimol, 318863743, yairb2@mail.tau.ac.il
- Ido Tamir, 322634213, idotamir1@mail.tau.ac.il
- Harel Ben Shoshan, 211786744, harelb2@mail.tau.ac.il

Project Description:

This project explores how the ordering of training data affects fact retention in small language models (LLMs). Specifically, we aim to test whether information presented at the beginning or end of a training dataset is more likely to be retained and retrieved accurately by a model. While transformer-based models process data in segments during training, it remains unclear whether global ordering across an entire corpus significantly impacts memorization or downstream performance.

To investigate this, we will train a small open-source LLM from scratch utilizing KAS. We will insert fictional biographical entries about individuals—each with a unique name and a small set of facts—at both the beginning and the end of the training corpus. After training, we will prompt the model about these individuals and compare accuracy and confidence across facts seen at different positions in the data.

If results are unclear or inconclusive, we will introduce contradicting facts about some individuals—placing different versions of a fact at the beginning and end of the dataset. We will then evaluate which version the model is more likely to generate, providing a stronger signal about the influence of data ordering.

This experiment will provide insight into the importance of dataset order in model training, particularly relevant for researchers optimizing limited training resources or curating datasets. We expect to find whether the early or late data is more influential and whether positional bias exists. Success will be measured by analyzing retrieval accuracy differences between the two groups, using standard evaluation techniques such as prompting and perplexity-based scoring.