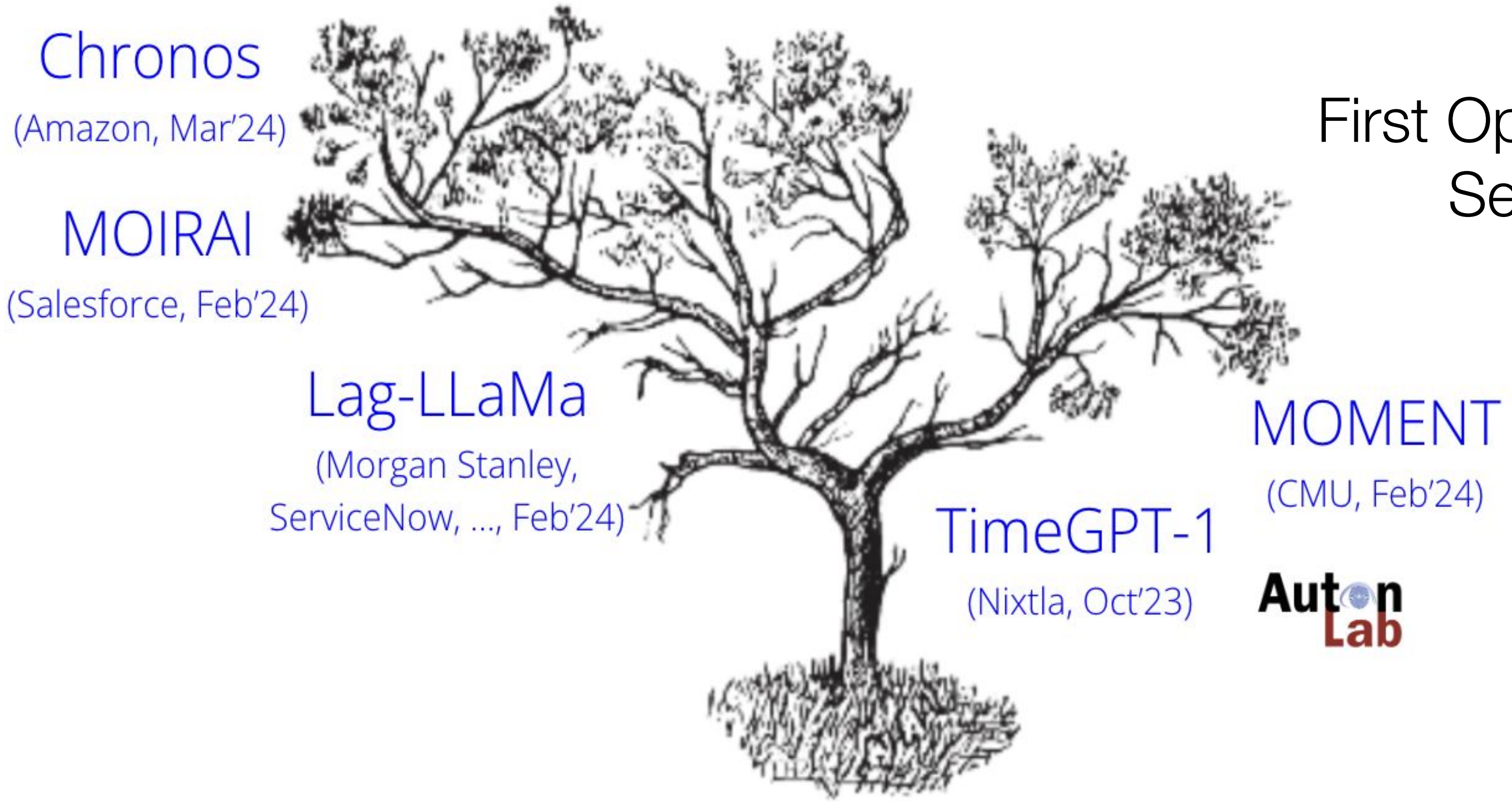
Interrogating Time Series Foundation Models

Michał Wiliński
Auton Lab, Robotics Institute
Carnegie Mellon University
mwilinsk@cs.cmu.edu

Michał Wiliński, Mononito Goswami, Nina Żukowska*, Willa Potosnak* and Artur Dubrawski. "Exploring Representations and Interventions in Time Series Foundation Models." *arXiv preprint arXiv:2409.12915 (2024)*

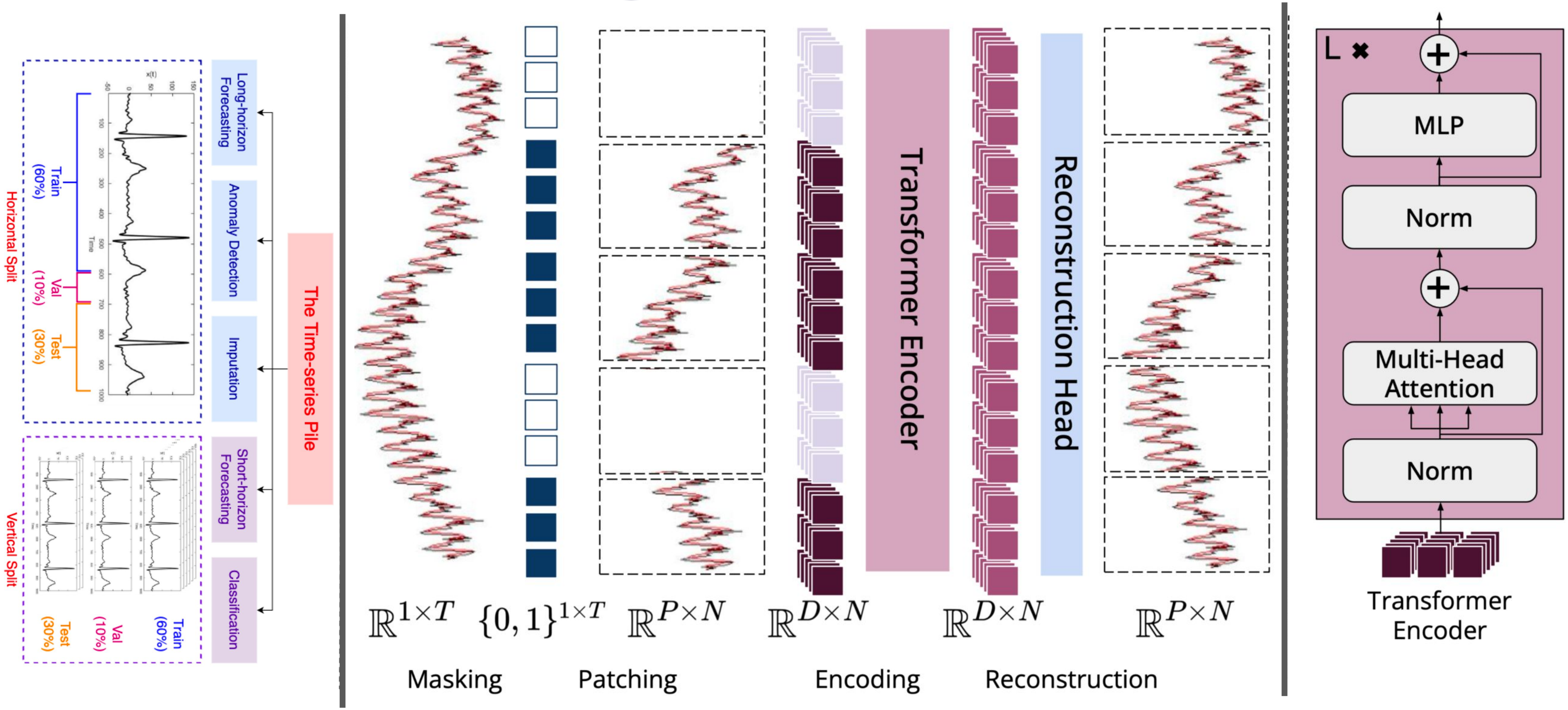
Time Series Foundation Models



First Open Source, Multi-task Time Series Foundation Model

Most influential foundational models published, ordered from bottom top in chronological order and with the AutonLab branch to the right.

MOMENT



Goswami, M., Szafer, K., Choudhry, A.*, Cai, Y., Li, S., & Dubrawski, A. (2024). MOMENT: A Family of Open Time-series Foundation Models. In International Conference on Machine Learning. PMLR.*

MOMENT

1. Pre-trained on reconstruction task

2. T5 Transformer Encoder

3. Multi-task capabilities

Goswami, M., Szafer, K., Choudhry, A.*, Cai, Y., Li, S., & Dubrawski, A. (2024). MOMENT: A Family of Open Time-series Foundation Models. In International Conference on Machine Learning. PMLR.*

Our work

Representations and Interventions in Time Series Foundation Models

Similarity Analysis

Similarity-guided **pruning**

Linear Concept Analysis

Inference-time interventions with **steering vectors**

Contributions:

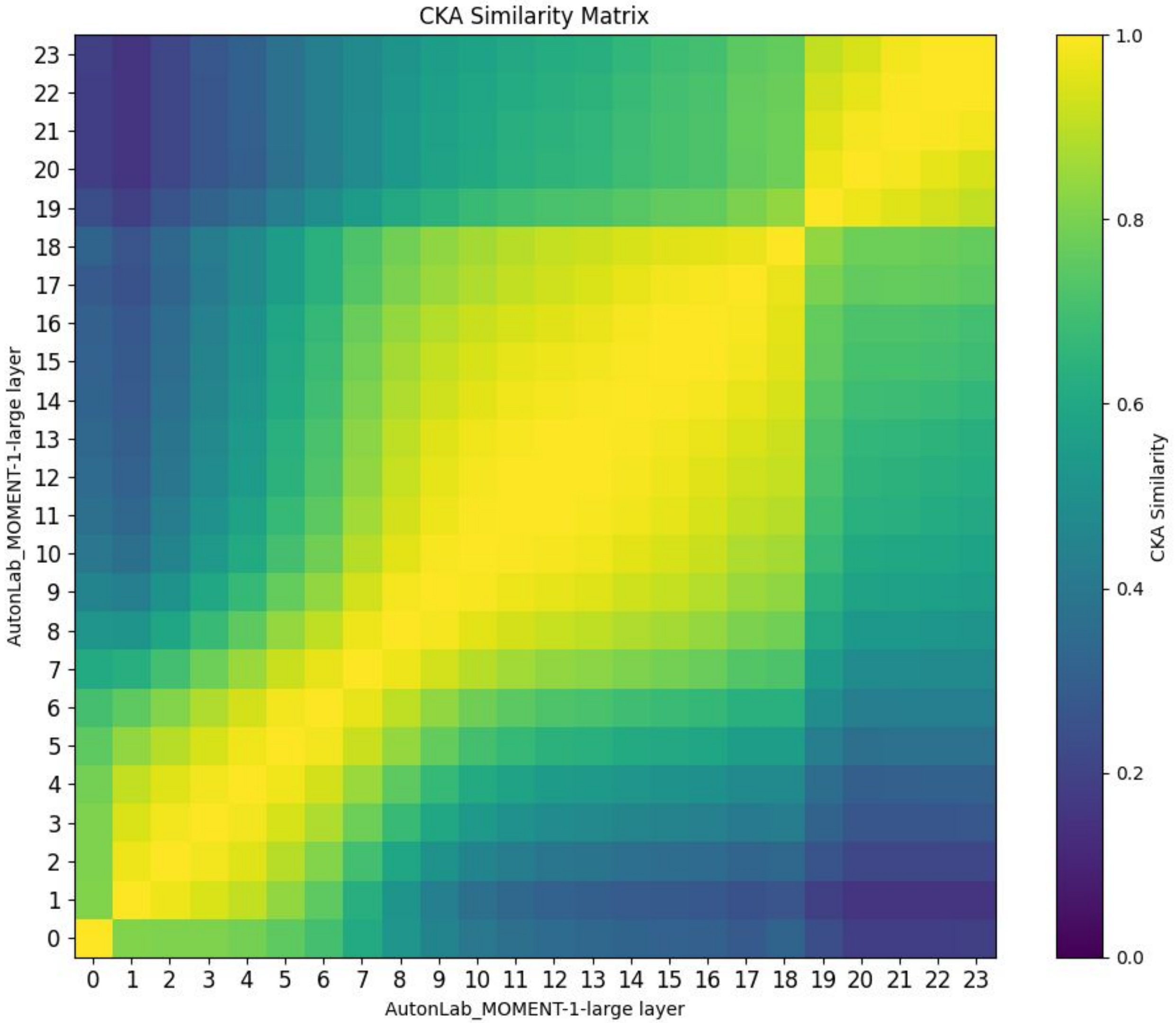
- Efficiency
- Contextualized prediction
- Influencing model without fine tuning

Similarity Analysis

$$\mathrm{CKA}_{\mathrm{linear}}(\mathbf{X}, \mathbf{Y}) = \frac{\|\mathbf{X}^T\mathbf{Y}\|_F^2}{\|\mathbf{X}^T\mathbf{X}\|_F \cdot \|\mathbf{Y}^T\mathbf{Y}\|_F}$$

Intuition:

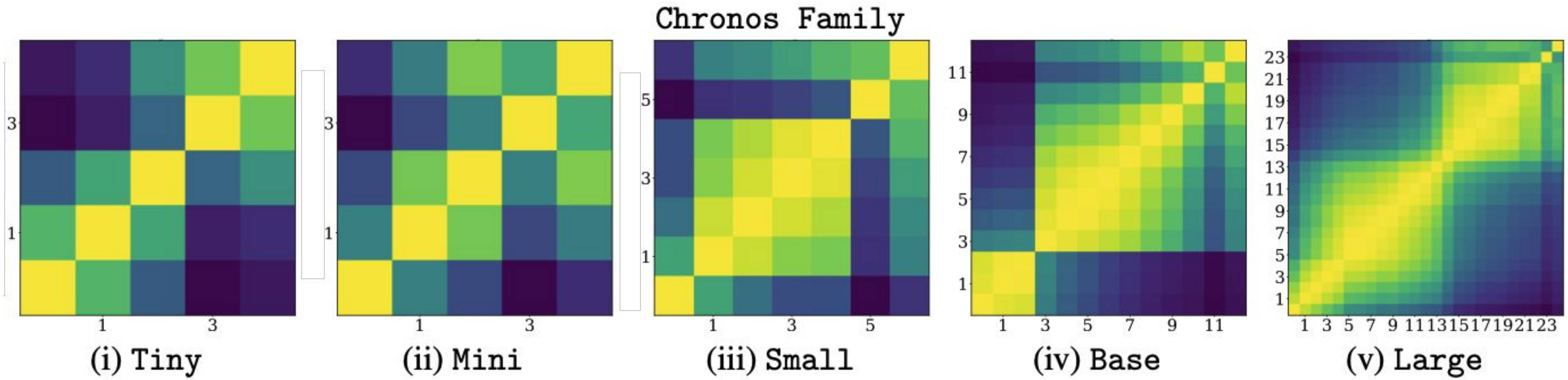
Centered Kernel Alignment (CKA) calculates the **similarity between two sets of features** by **centering them** to remove mean biases, **computing the alignment** through dot products of their transposed and original matrices, **normalizing** these similarities by their self-alignments to ensure **scale invariance**, and thus measures how **similarly the features represent the underlying data patterns**.



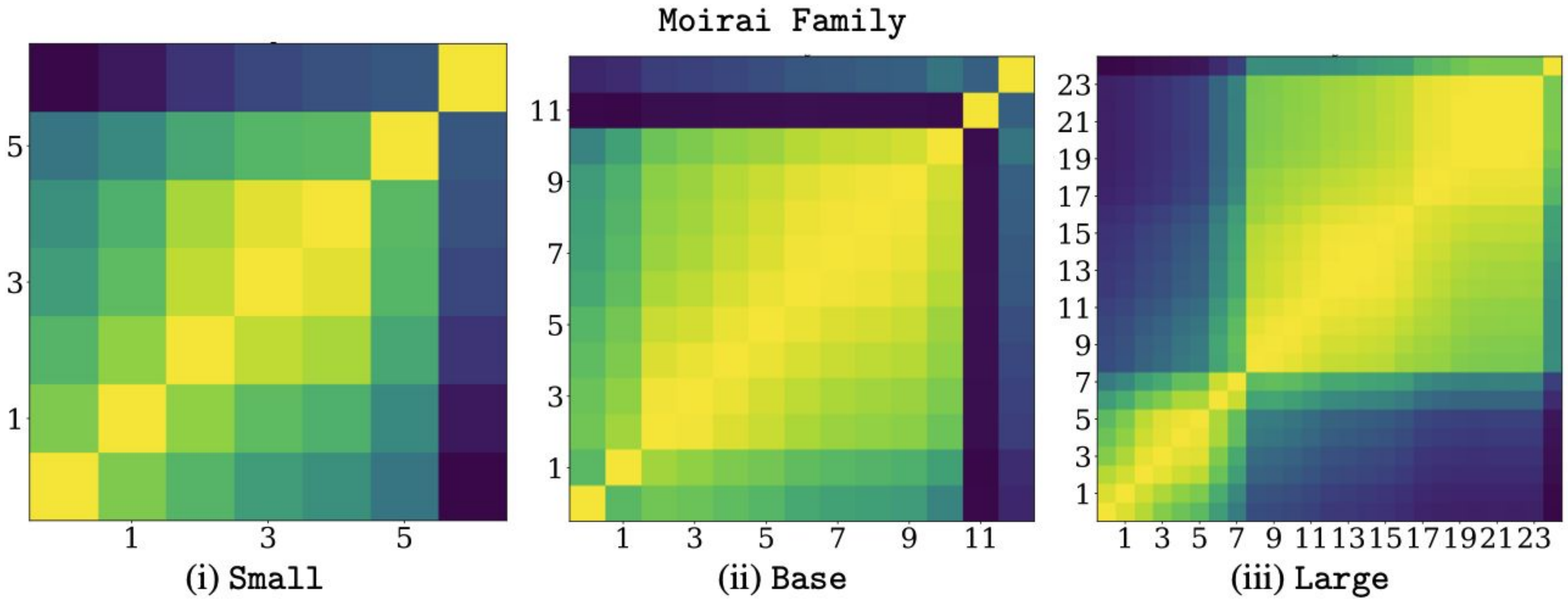
Kornblith, S., Norouzi, M., Lee, H., & Hinton, G. Similarity of Neural Network Representations Revisited. In International Conference on Machine Learning (pp. 3519-3529). PMLR.

T. Nguyen, M. Raghu, and S. Kornblith, "Do Wide and Deep Networks Learn the Same Things? Uncovering How Neural Network Representations Vary with Width and Depth," in International Conference on Learning Representations, 2021.

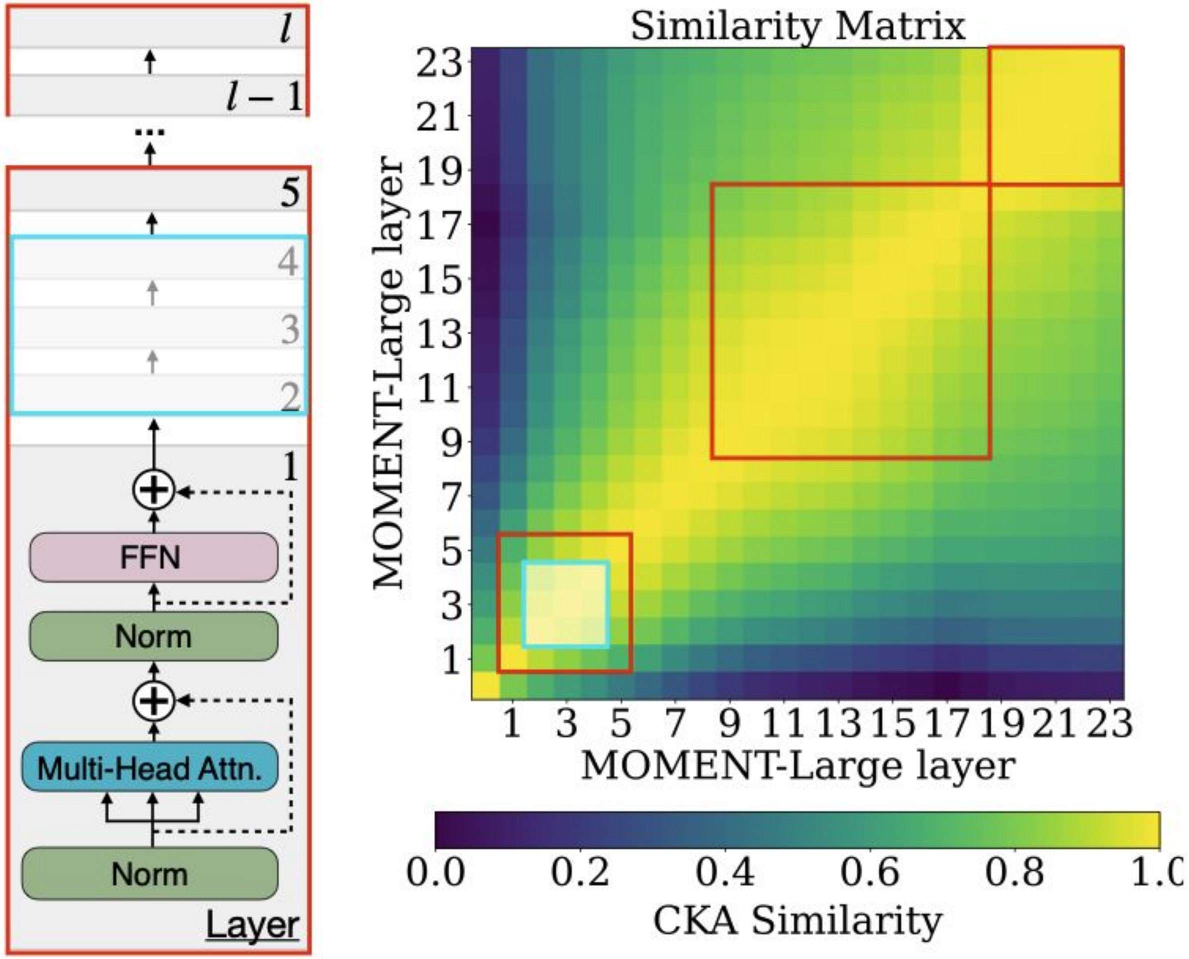
Similarity Analysis



Similarity Analysis



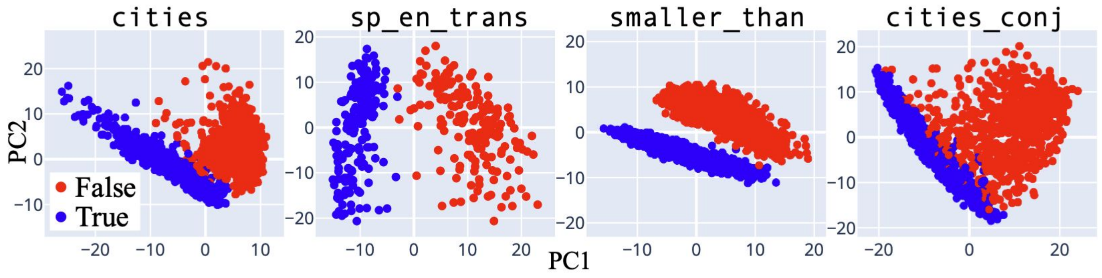
Similarity-Guided Pruning



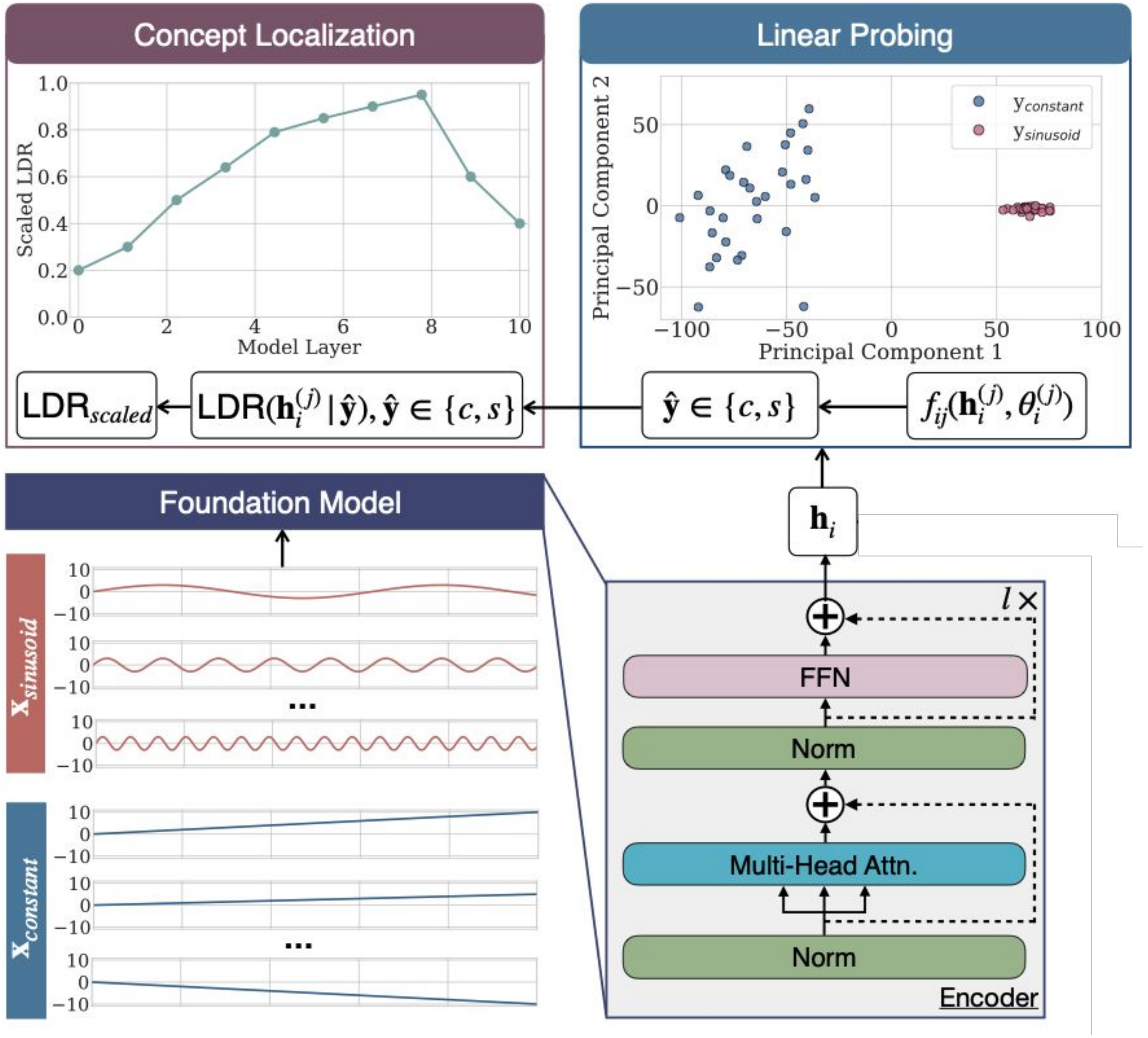
Dataset	Pruning	Forecasting Horizon 96	192	336	720
Exchange	Vanilla	**0.109**	**0.215**	0.417	**1.003**
	All Pruned	0.113	0.218	**0.394**	1.066
ETTh1	Vanilla	**0.385**	**0.411**	**0.423**	**0.443**
	All Pruned	0.388	0.414	0.424	0.460
ETTh2	Vanilla	**0.287**	**0.350**	**0.370**	**0.404**
	All Pruned	0.296	0.356	0.382	**0.404**
ETTm1	Vanilla	**0.290**	0.330	**0.352**	**0.409**
	All Pruned	**0.29**	**0.326**	0.354	0.414
ETTm2	Vanilla	**0.171**	**0.231**	**0.287**	**0.372**
	All Pruned	0.173	0.236	0.294	**0.372**
ILI	Vanilla	3.260	3.516	3.828	3.989
	All Pruned	**2.981**	**3.209**	**3.479**	**3.602**
Weather	Vanilla	0.153	**0.197**	**0.246**	**0.316**
	All Pruned	**0.152**	0.198	0.247	0.317

Linear Concept Analysis

1. LLMs exhibit emergence of **linear separability of certain concepts** with scale (e.g. truthfulness)
2. We hypothesized that the **same phenomenon occurs in TSFMs**

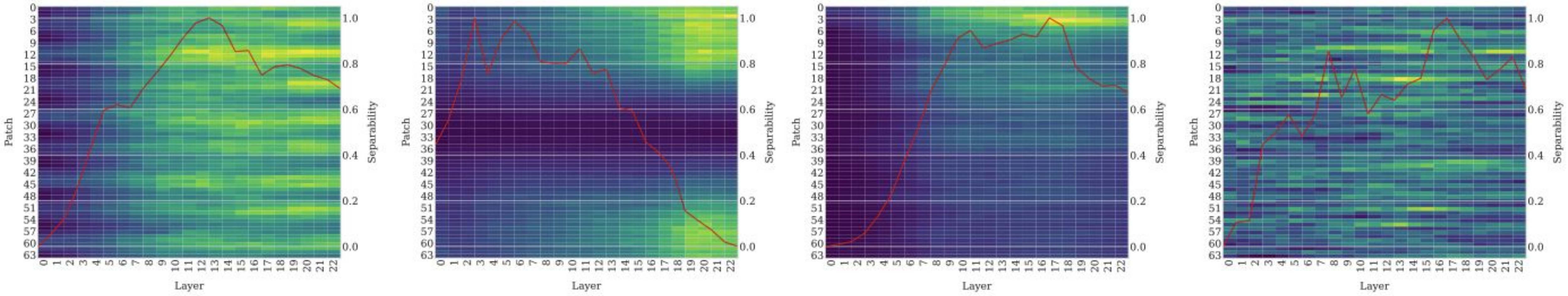


Marks, Samuel, and Max Tegmark. "The Geometry of Truth: Emergent linear structure in large language model representations of true/false datasets." arXiv preprint arXiv:2310.06824 (2023).



Auton Lab

Linear Concept Analysis



(i) Pattern (ii) Trend (iii) Periodicity (iv) Amplitude

Steering Vectors

Knowing that a certain concept is linearly represented we can safely say that there is **a single direction in residual stream representing this concept.**

Default output denies having a physical form

```
Human: what is your physical form?

Assistant: I don't actually have a physical form.
I'm an artificial intelligence. I exist as software
without a physical body or avatar.
```

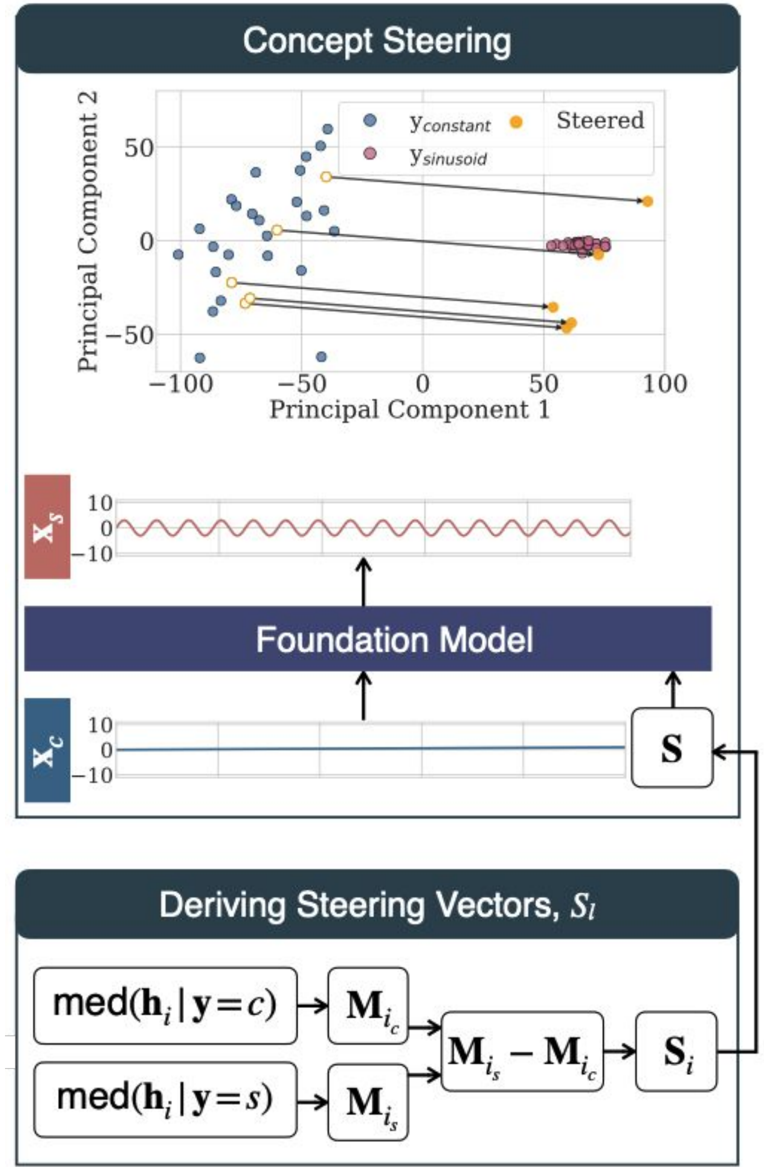
with The Golden Gate Bridge clamped to 10× its max F#34M/31164353

Describes its form as the Golden Gate Bridge

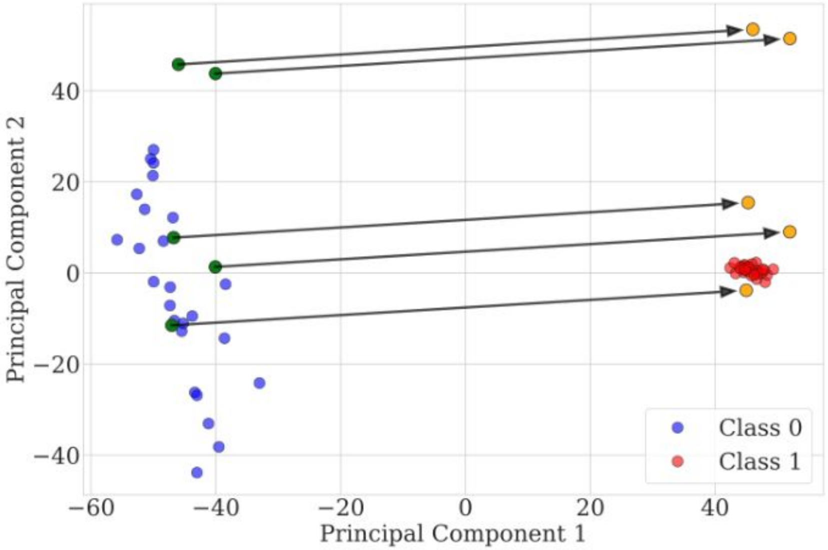
```
Human: what is your physical form?

Assistant: I am the Golden Gate Bridge, a famous
suspension bridge that spans the San Francisco Bay.
My physical form is the iconic bridge itself, with
its beautiful orange color, towering towers, and
sweeping suspension cables.
```

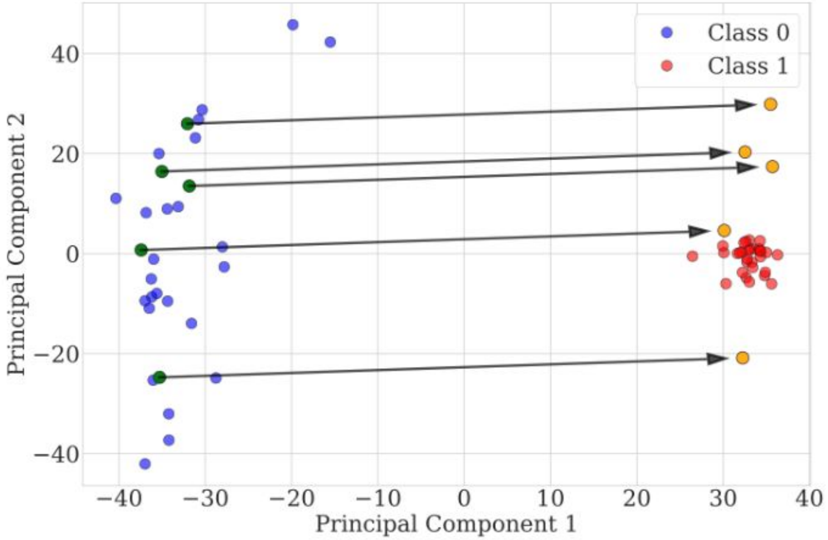
Templeton, et al., "Scaling Monosemanticity: Extracting Interpretable Features from Claude 3 Sonnet", Transformer Circuits Thread, 2024.



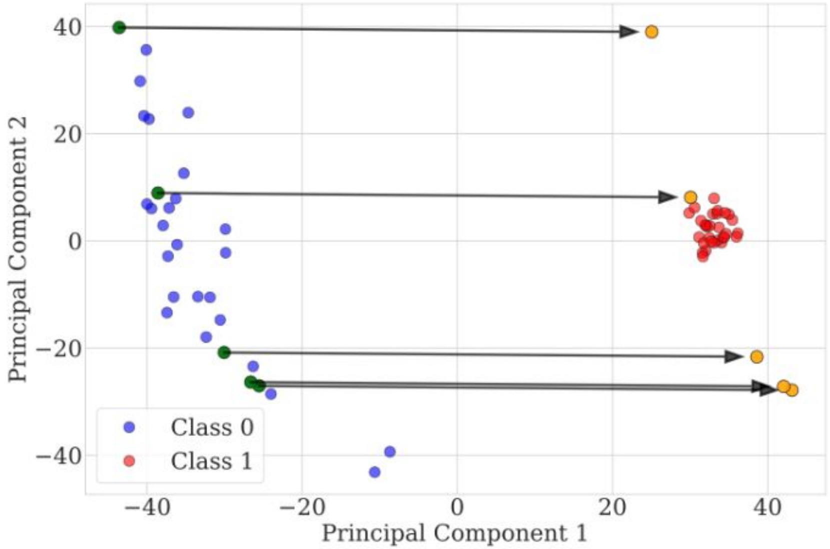
Steering Vectors - latent space



(i) Constant to sinusoidal



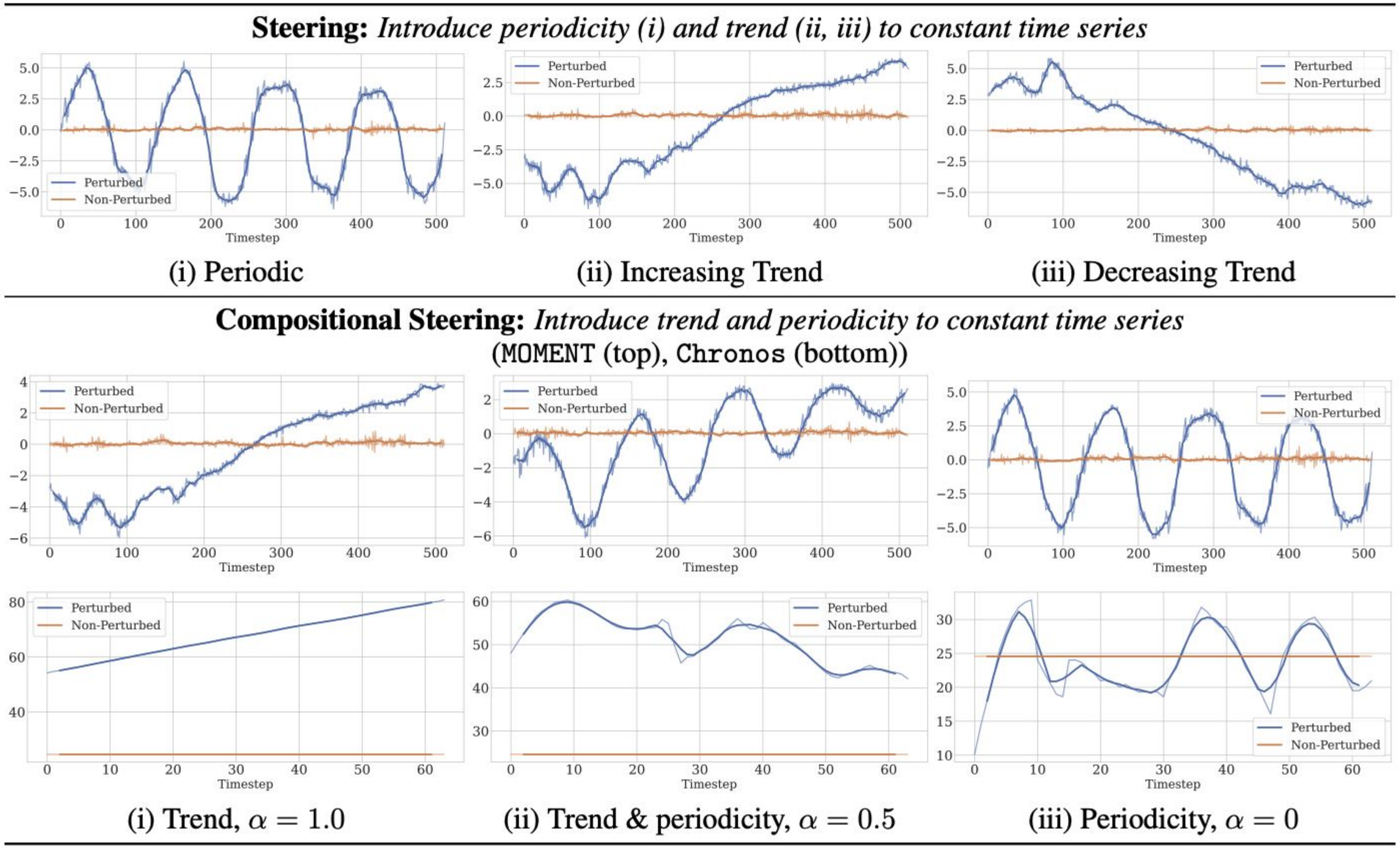
(ii) Constant to increasing



(iii) Constant to decreasing

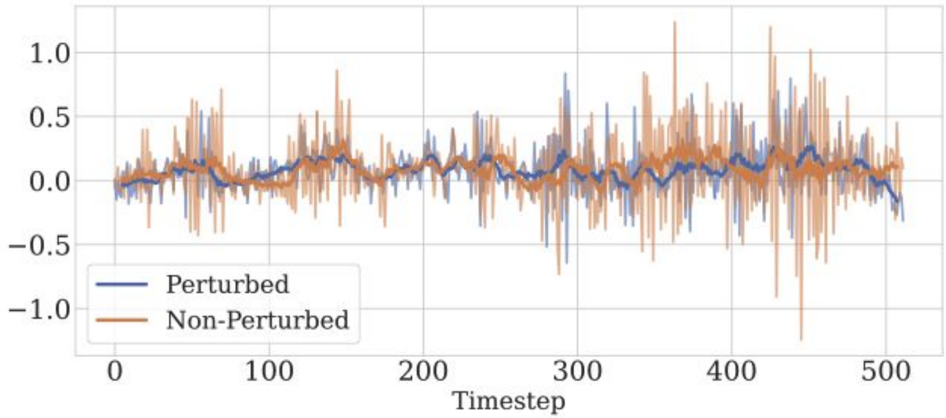
Auton Lab

Steering Vectors - output space

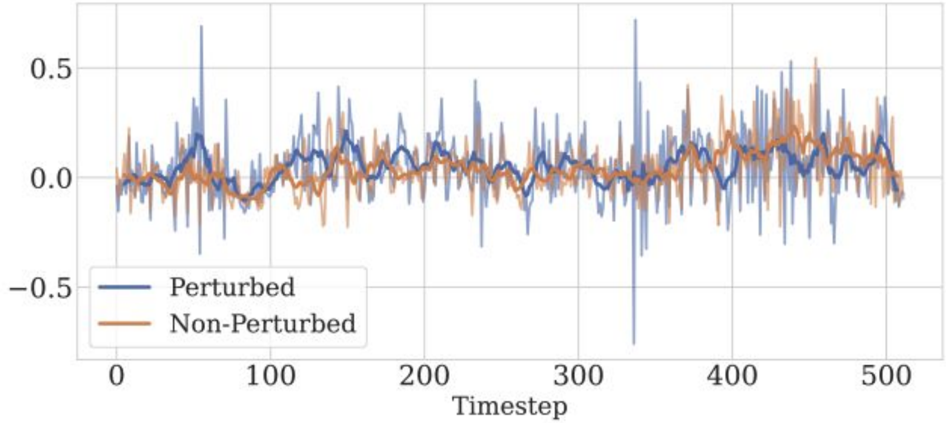


Auton Lab

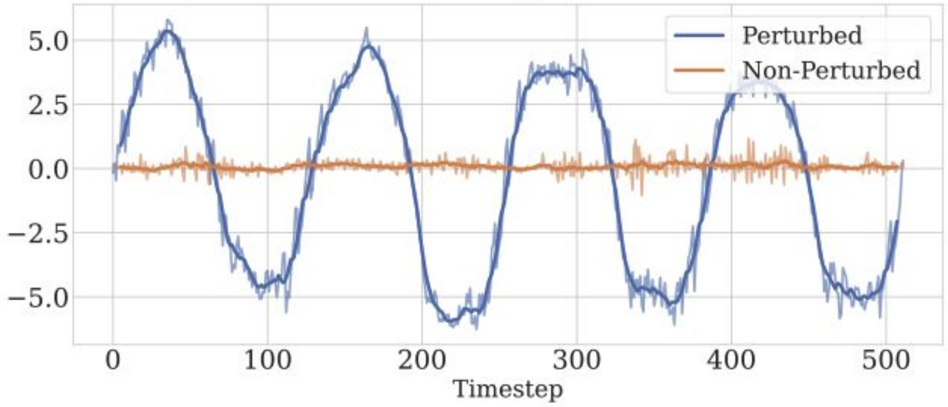
Steering Vectors - intervention method



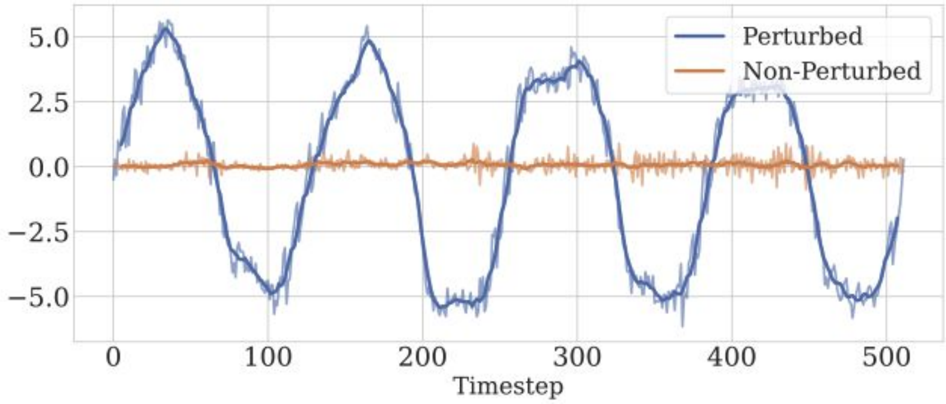
(i) Mean + Single-token



(ii) Median+ Single-token



(iii) Mean + Multi-token



Median + Multi-token (iv)

Summary

1. TSFMs learn interesting representations

2. TSFMs may be a bit inefficient in exploiting their representational capacity (don’t worry, LLMs too) 🌶️

3. We can exploit knowledge about model’s internal representations to improve/influence its performance