Geometric Deep Learning

14. Applications II

Jun.-Prof. Dr. Jan Stühmer

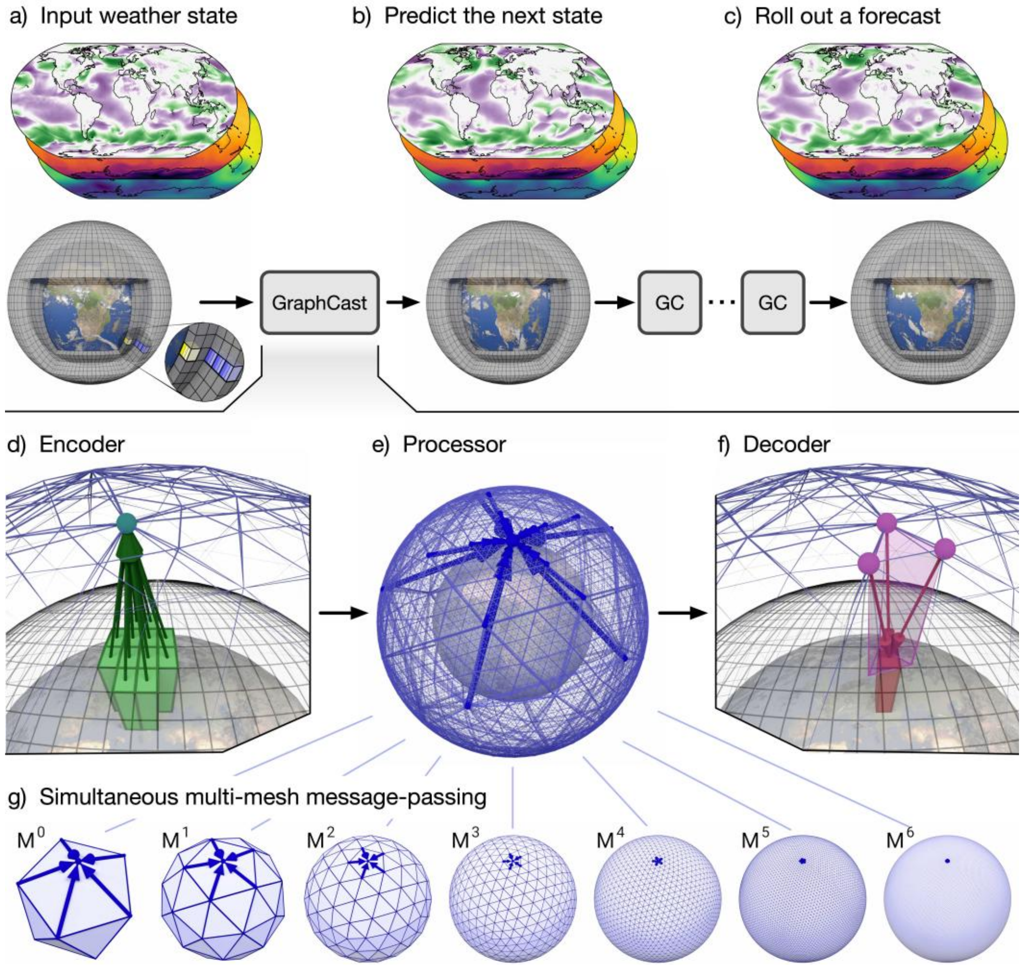
Outline

- Weather prediction
  - Multiresolution GNN: GraphCast [Lam et al. 2022]
  - Fourier neural operators: FourCastNet [Kurth et al. 2022]
  - + Spherical harmonics: FourCastNet 3 [Bonev et al. 2025]
- Set-LLM: A permutation equivariant transformer

Multi-Resolution GNN: GraphCast

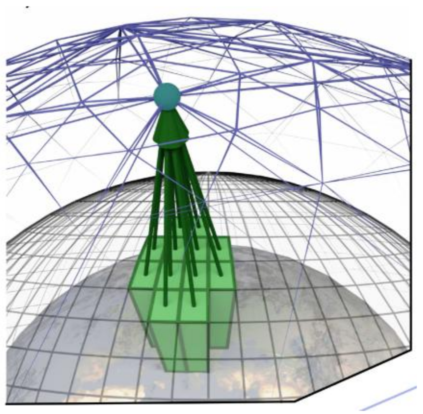
GraphCast [Lam et al. 2022]

- GraphCast implements multi-resolution principles in a graph neural network architecture
- Equiviariance: Translation group by weight sharing and permutation group of the graph node indices

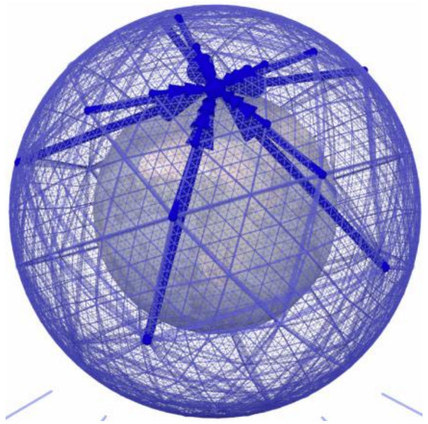


Multi-Resolution Graph Neural Network

- Grid-to-Mesh (G2M) Message-Passing-GNN
  - Edge update: $$\mathbf{e}^{\mathrm{G2M}\,\prime}_{v^{\mathrm{G}}_{\mathrm{s}}\to v^{M}_{\mathrm{r}}} = \mathrm{MLP}^{\mathrm{Grid2Mesh}}_{\mathcal{E}^{\mathrm{G2M}}}\left(\left[\mathbf{e}^{\mathrm{G2M}}_{v^{\mathrm{G}}_{\mathrm{s}}\to v^{M}_{\mathrm{r}}}, \mathbf{v}^{\mathrm{G}}_{s}, \mathbf{v}^{M}_{r}\right]\right)$$
  - Node update: $$\mathbf{v}^{\mathrm{M}\prime}_{i} = \mathrm{MLP}^{\mathrm{Grid2Mesh}}_{\mathcal{V}^{\mathrm{M}}}\left(\left[\mathbf{v}^{\mathrm{M}}_{i}, \sum_{e^{\mathrm{G2M}}_{v^{\mathrm{G}}_{\mathrm{s}}\to v^{\mathrm{M}}_{\mathrm{r}}}\,:\, v^{\mathrm{M}}_{\mathrm{r}}=v^{\mathrm{M}}_{i}} \mathbf{e}^{\mathrm{G2M}\,\prime}_{v^{\mathrm{G}}_{\mathrm{s}}\to v^{\mathrm{M}}_{\mathrm{r}}}\right]\right)$$

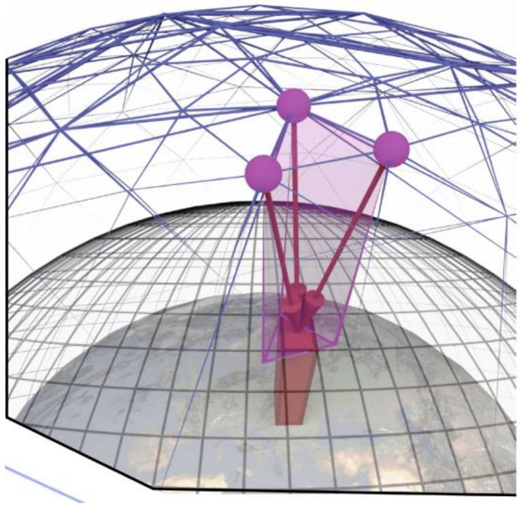


- Mesh-to-Mesh Message-Passing-GNN:
  - Edge update: $$\mathbf{e}^{\mathrm{M}\,\prime}_{v^{M}_{\mathrm{s}}\to v^{M}_{\mathrm{r}}} = \mathrm{MLP}^{\mathrm{Mesh}}_{\mathcal{E}^{\mathrm{M}}}\left(\left[\mathbf{e}^{\mathrm{M}}_{v^{M}_{\mathrm{s}}\to v^{M}_{\mathrm{r}}}, \mathbf{v}^{M}_{s}, \mathbf{v}^{M}_{r}\right]\right)$$
  - Node update: $$\mathbf{v}^{\mathrm{M}\prime}_{i} = \mathrm{MLP}^{\mathrm{Mesh}}_{\mathcal{V}^{\mathrm{M}}}\left(\left[\mathbf{v}^{\mathrm{M}}_{i}, \sum_{e^{\mathrm{M}}_{v^{\mathrm{M}}_{\mathrm{s}}\to v^{\mathrm{M}}_{\mathrm{r}}}\,:\, v^{\mathrm{M}}_{\mathrm{r}}=v^{\mathrm{M}}_{i}} \mathbf{e}^{\mathrm{M}\,\prime}_{v^{\mathrm{M}}_{\mathrm{s}}\to v^{\mathrm{M}}_{\mathrm{r}}}\right]\right)$$



Multi-Resolution Graph Neural Network

- Mesh-to-Grid (M2G) Message-Passing-GNN
  - Edge update: $$\mathbf{e}^{\mathrm{M2G}}_{v^M_{\mathrm{s}} \to v^{\mathrm{G}}_{\mathrm{r}}}{}' = \mathrm{MLP}^{\mathrm{Mesh2Grid}}_{\mathcal{E}^{\mathrm{M2G}}}\left([\mathbf{e}^{\mathrm{M2G}}_{v^M_{\mathrm{s}} \to v^{\mathrm{G}}_{\mathrm{r}}}, \mathbf{v}^M_s, \mathbf{v}^{\mathrm{G}}_r]\right)$$
  - Node update: $$\mathbf{v}^{\mathrm{G}\prime}_i = \mathrm{MLP}^{\mathrm{Mesh2Grid}}_{\mathcal{V}^{\mathrm{G}}}\left([\mathbf{v}^{\mathrm{G}}_i, \sum_{e^{\mathrm{M2G}}_{v^{\mathrm{M}}_{\mathrm{s}} \to v^{\mathrm{G}}_{\mathrm{r}}} :\, v^{\mathrm{G}}_{\mathrm{r}} = v^{\mathrm{G}}_i} \mathbf{e}^{\mathrm{M2G}}_{v^{\mathrm{M}}_{\mathrm{s}} \to v^{\mathrm{G}}_{\mathrm{r}}}{}'\right])$$



- Output function per grid node:

$$\hat{\mathbf{y}}^{\mathrm{G}}_i = \mathrm{MLP}^{\mathrm{Output}}_{\mathcal{V}^{\mathrm{G}}}(\mathbf{v}^{\mathrm{G}}_i)$$

Fourier Neural Operators: FourCastNet

Adaptive Fourier Neural Operators [Guibas et al 2022]

Fourier Neural Operator:

$$\mathcal{K}(X)(u) = \mathcal{F}^{-1}\big(\mathcal{F}(\kappa) \cdot \mathcal{F}(X)\big)(u)$$

for a continuous input X and kernel κ. $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the continuous Fourier transform and its inverse.

Compare to spectral convolution (Lecture 11):

$$(x * \psi)(u) = \sum_{k \geq 0} \left(\hat{x}_k \cdot \hat{\psi}_k\right) \varphi_k(u)$$

Adaptive Fourier Neural Operators [Guibas et al 2022]

Discrete Fourier Neural Operator:

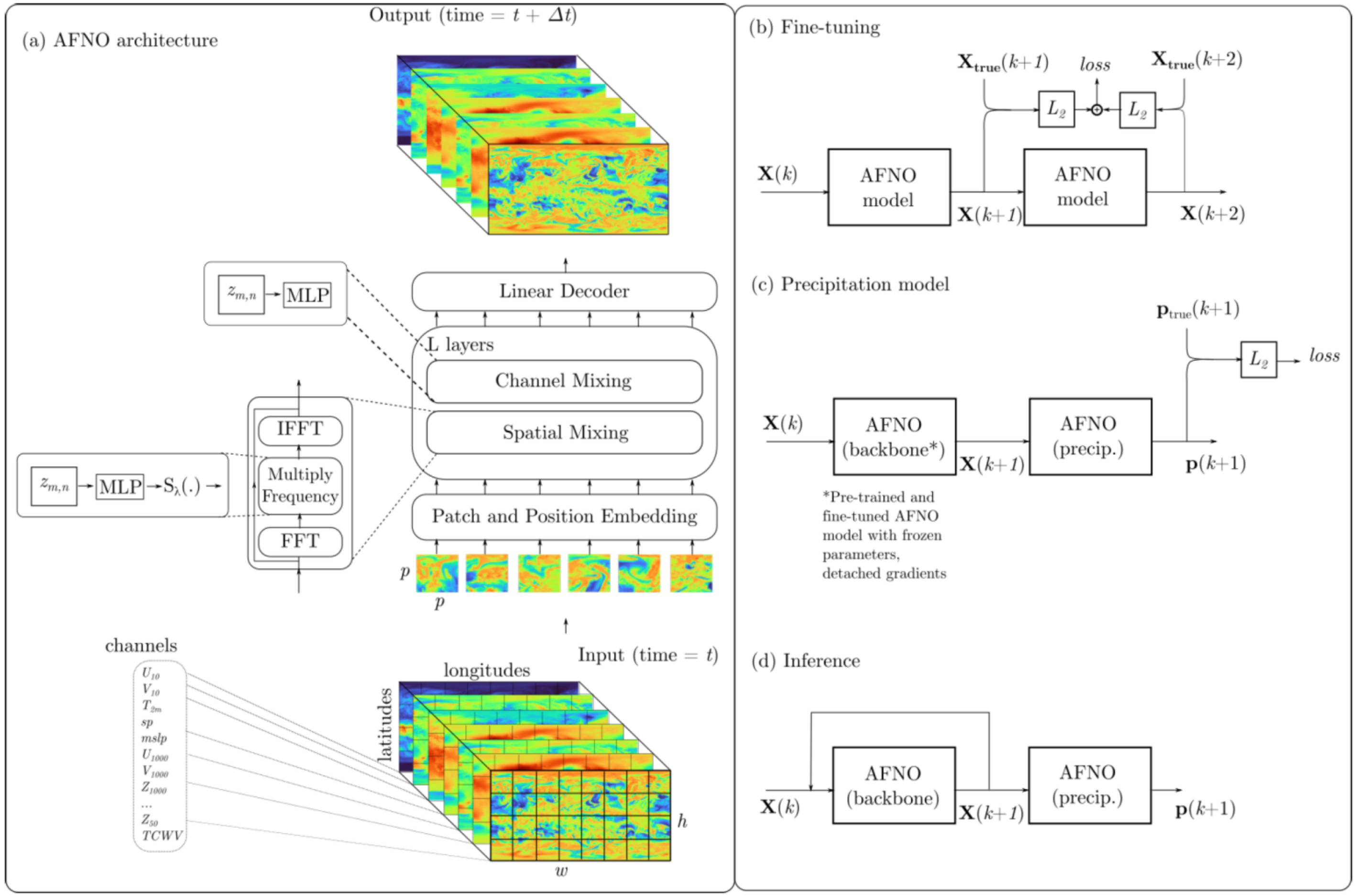
step (1). token mixing $$z_{m,n} = [\mathrm{DFT}(X)]_{m,n}$$

step (2). channel mixing $$\tilde{z}_{m,n} = W_{m,n} z_{m,n}$$

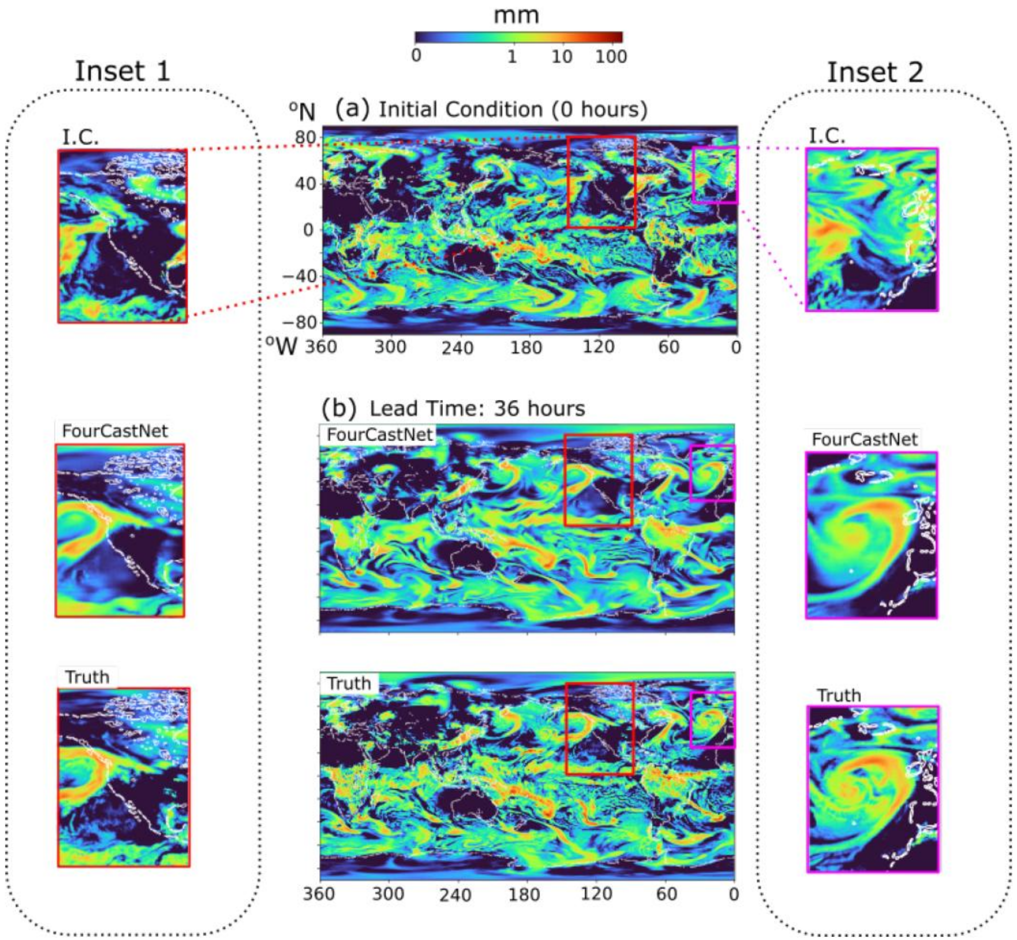
step (3). token demixing $$y_{m,n} = [\mathrm{IDFT}(\tilde{Z})]_{m,n}$$

Which amounts to a spectral convolution with channel mixing.

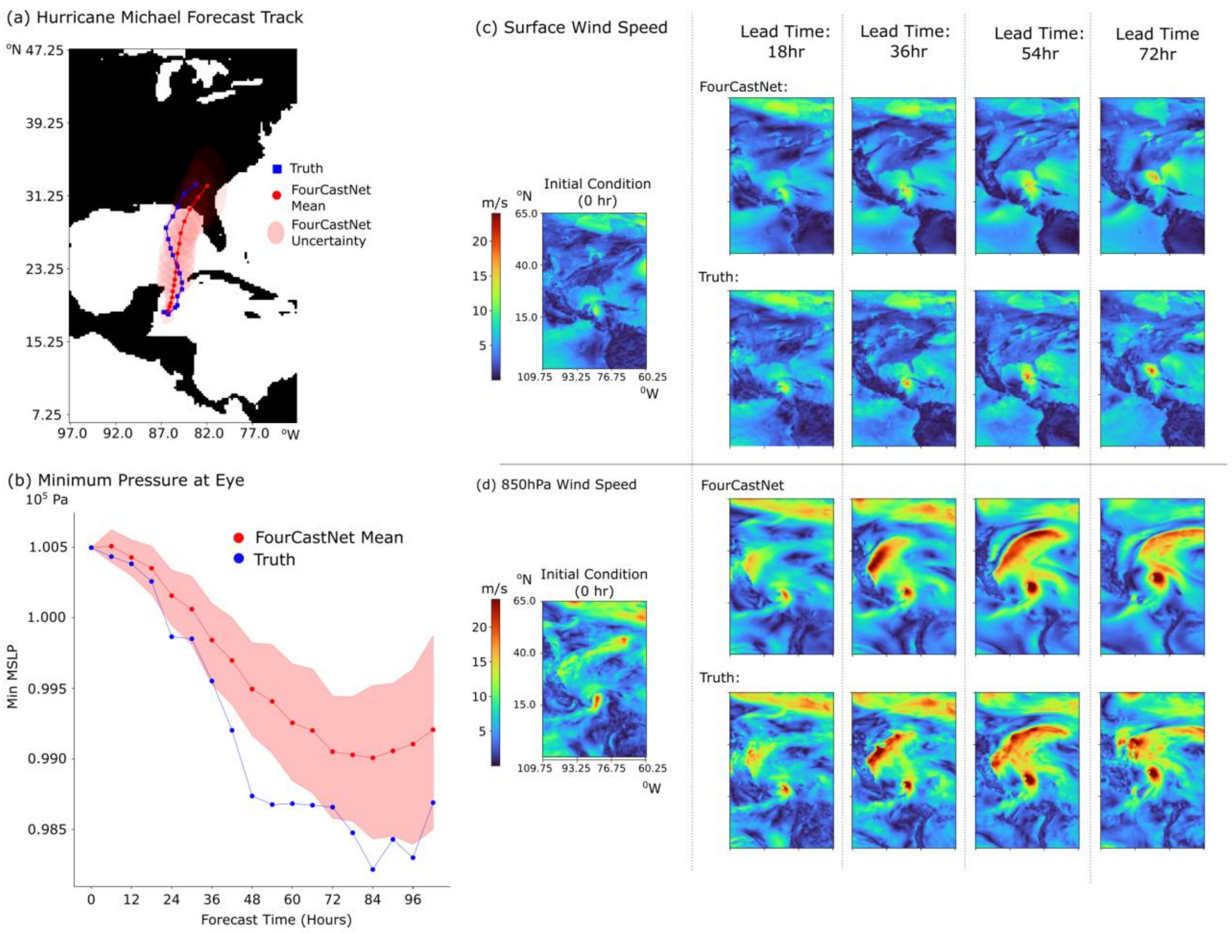
FourCastNet [Kurth et al. 2022] - Architecture



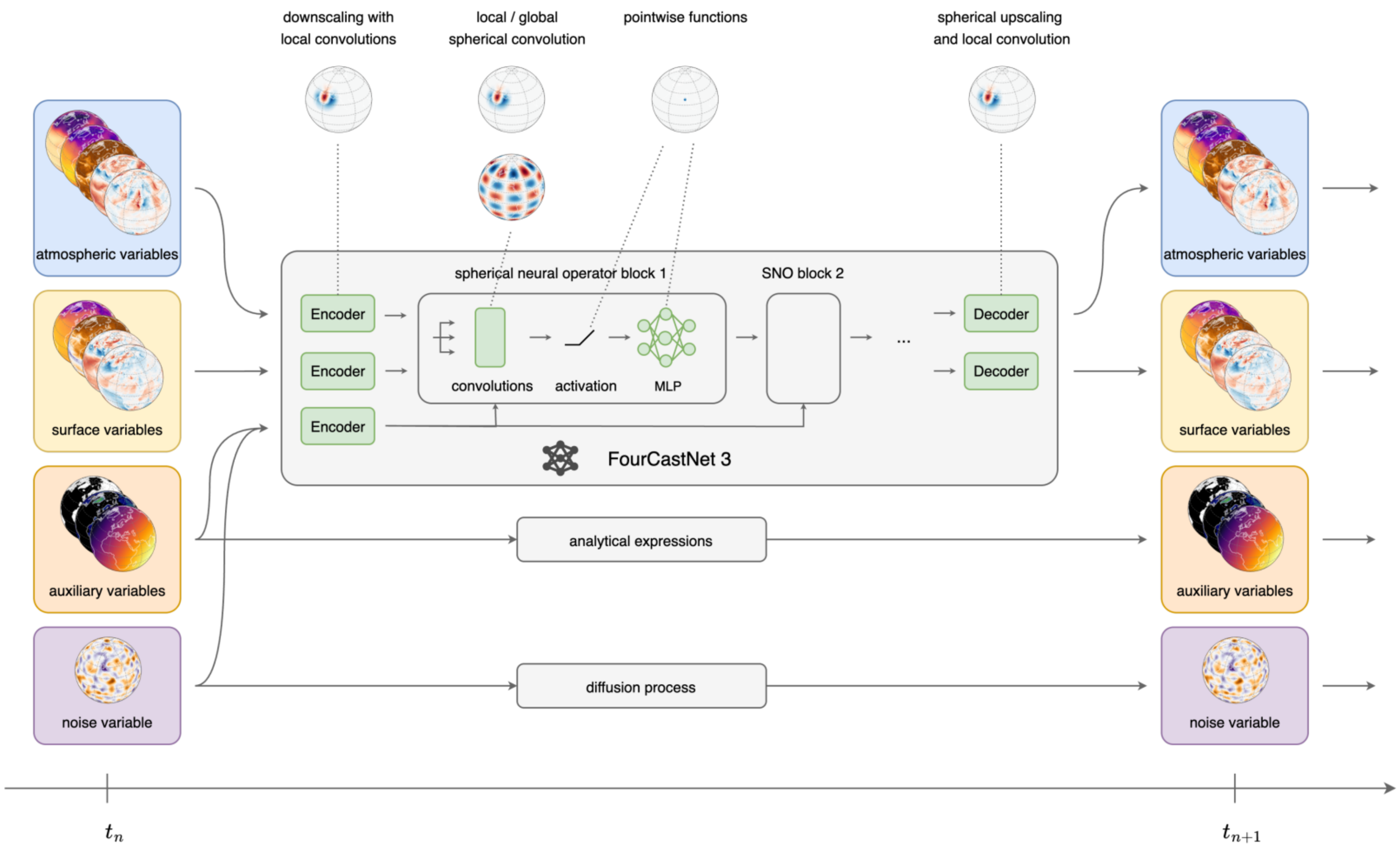
FourCastNet – Precipitation due to Cyclone



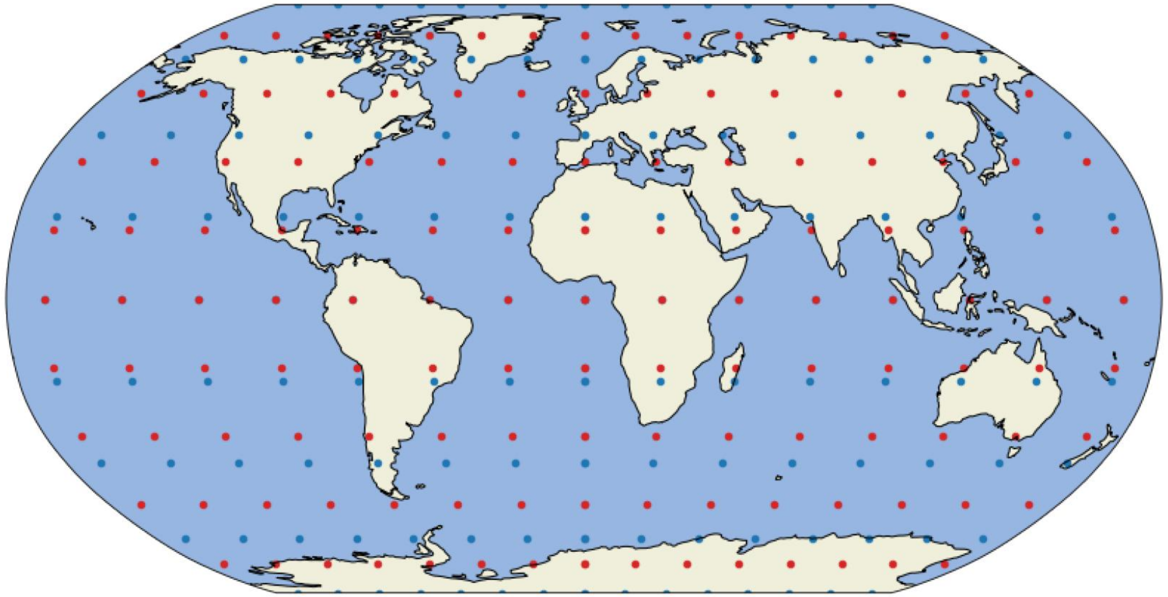
FourCastNet – Hurricane Forecast



FourCastNet 3

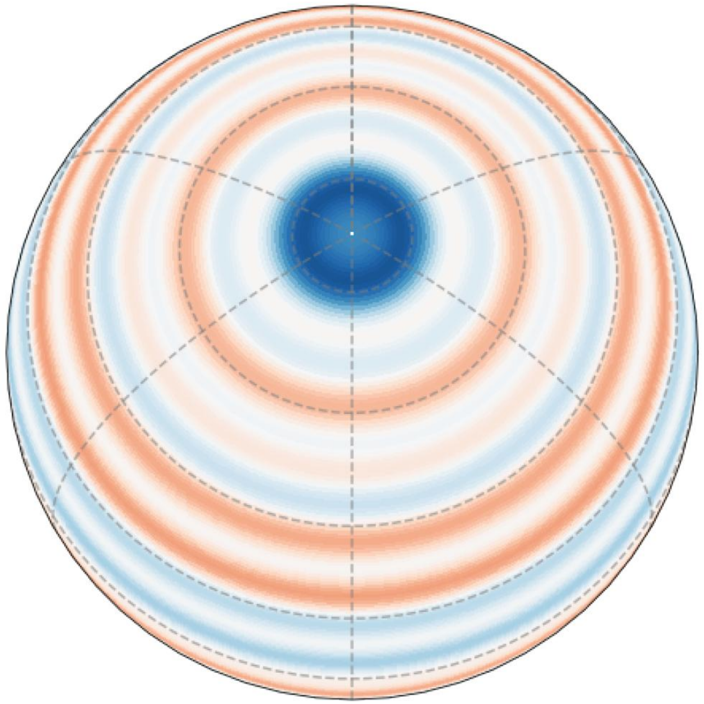


Convolutions on the Sphere

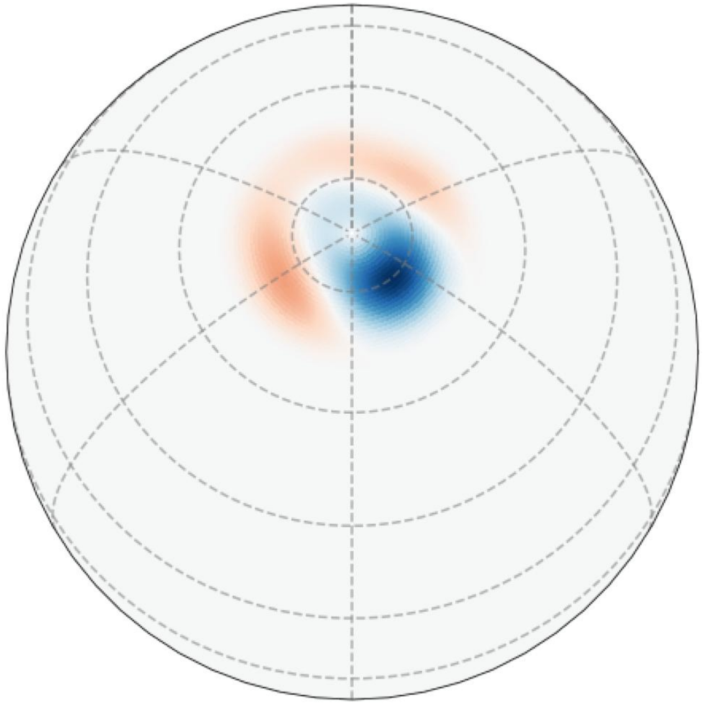


Compare also to Spherical CNNs [Cohen et al. 2018], Clebsch-Gordan Nets [Kondor et al. 2018], Discrete-Continuous-Convolutions (DISCO) [Ocampo et al. 2023]

Spherical Convolutions



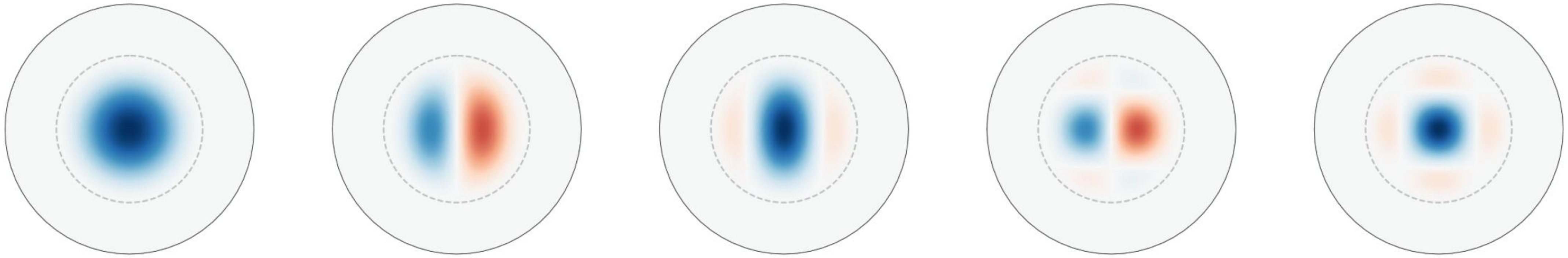
Spectral convolution filter
(linear combination of globally supported spherical harmonics basis functions)



Local convolution filter
(linear combination of locally supported spherical harmonics wavelets)

Wavelet Filter on the Sphere

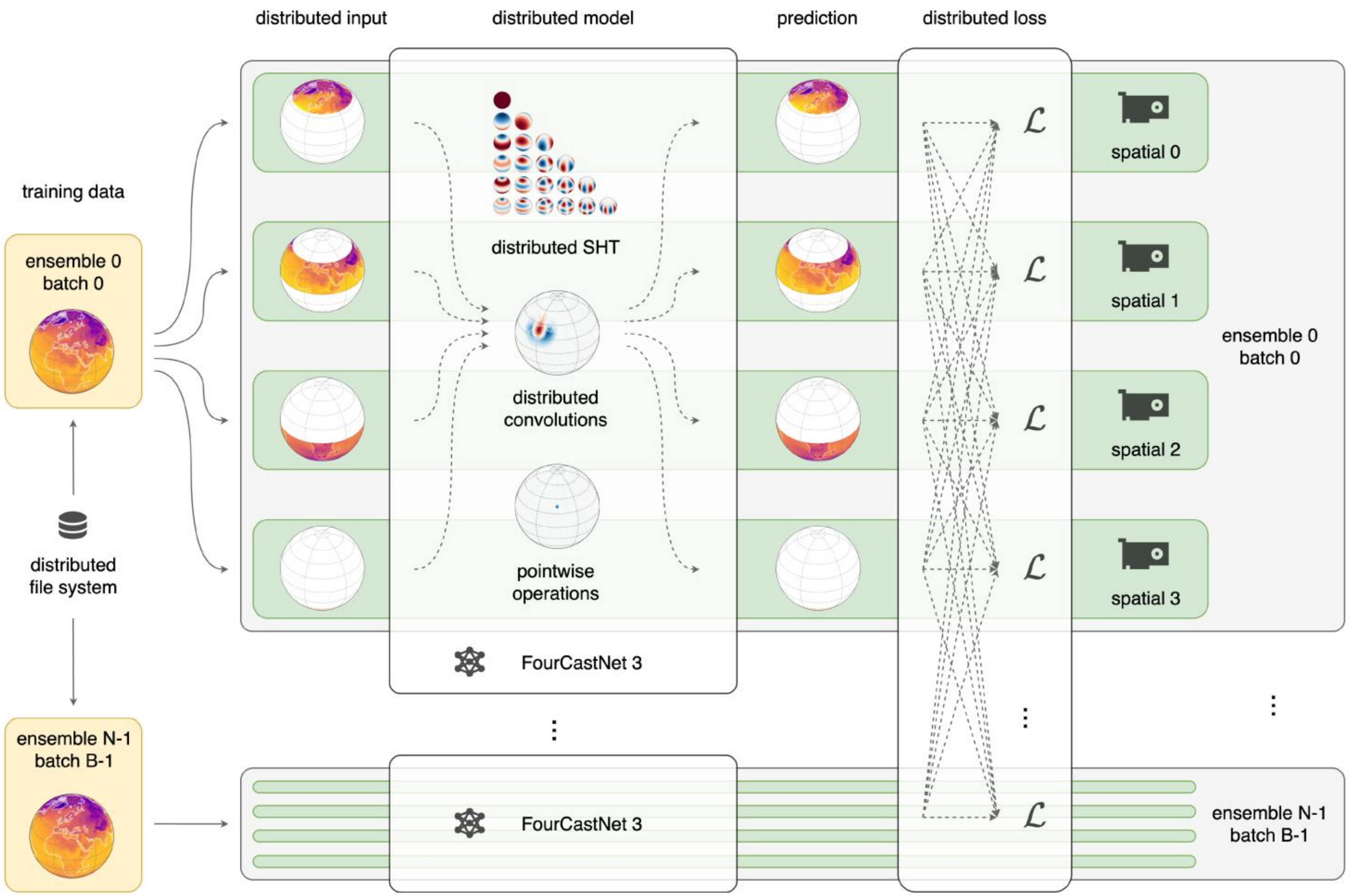
- The wavelet filter basis represents localized filters on the sphere:



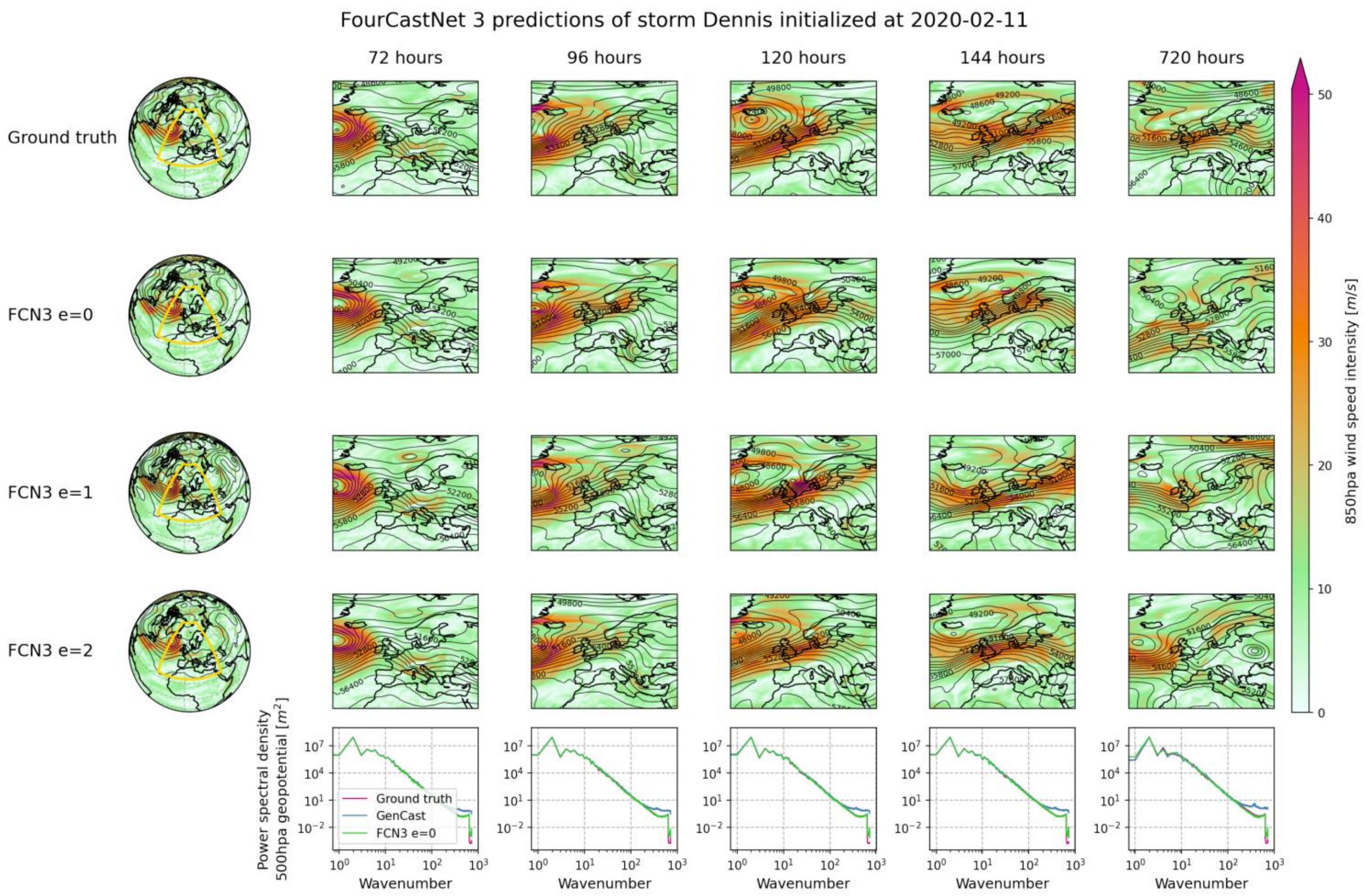
(a) $\ell = 0, m = 0$ (b) $\ell = 0, m = 1$ (c) $\ell = 0, m = 2$ (d) $\ell = 2, m = 1$ (e) $\ell = 2, m = 2$

Morlet-type wavelet filter basis. The dotted line represents the cutoff radius.

FourCastNet 3 – Data Parallelism



FourCastNet 3 – Ensemble Predictions



Further Applications

- Traffic forecasting:
  - Arrival time prediction [Derrow-Pinion et al., 2021]

- Particle Physics:
  - Lorentz-Equivariant Models [Bogatskiy et al., 2020, Brehmer et al. 2024]

Conclusion

- We have discussed several models for weather prediction that incorporate ideas of the geometric deep learning blueprint:
  - Multi-Resolution with a GNN: GraphCast
  - Fourier-Neural-Operator / spectral convolution with channel mixing: FourCastNet
  - SO(3) Equivariance with spherical harmonics and localization through wavelets: FourCastNet 3

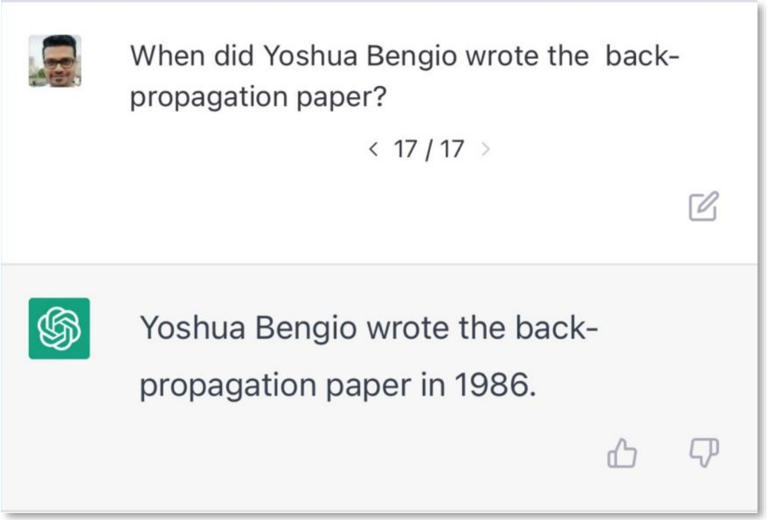
References

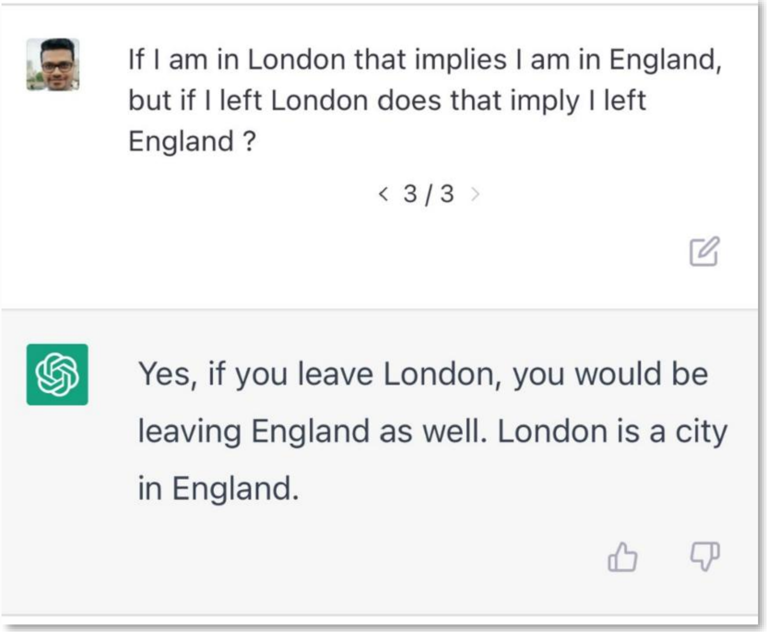
- GraphCast
  - Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., ... & Battaglia, P. Learning skillful medium-range global weather forecasting. *Science*, *382*(6677), 1416-1421. 2023.
- FourCastNet
  - Kurth, T., Subramanian, S., Harrington, P., Pathak, J., Mardani, M., Hall, D., ... & Anandkumar, A. Fourcastnet: Accelerating global high-resolution weather forecasting using adaptive fourier neural operators. In *Proceedings of the platform for advanced scientific computing conference* (pp. 1-11). 2023.
- FourCastNet 3
  - Bonev, Boris, et al. "Fourcastnet 3: A geometric approach to probabilistic machine-learning weather forecasting at scale." arXiv preprint arXiv:2507.12144. 2025.

Set-LLM: A permutation invariant LLM

with Beni Egressy

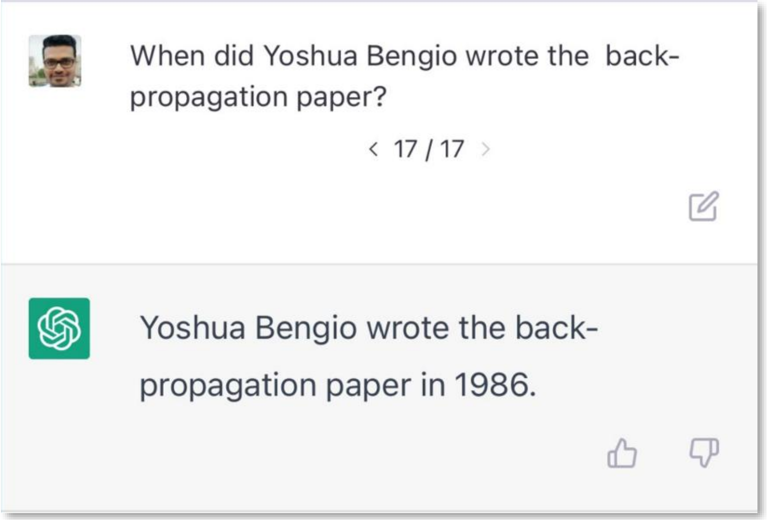
„Hallucination“ in Language Models

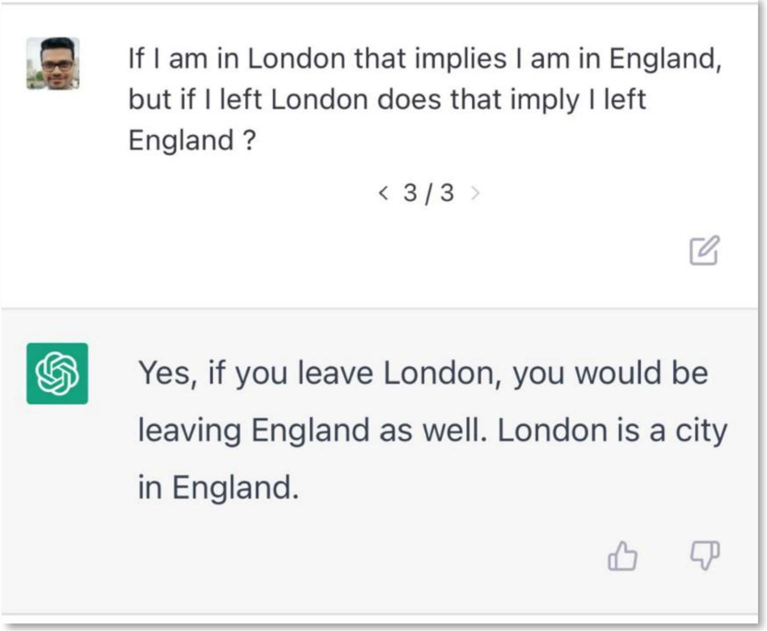




[Prithivi Da, via LinkedIn]

„Hallucination“ in Language Models

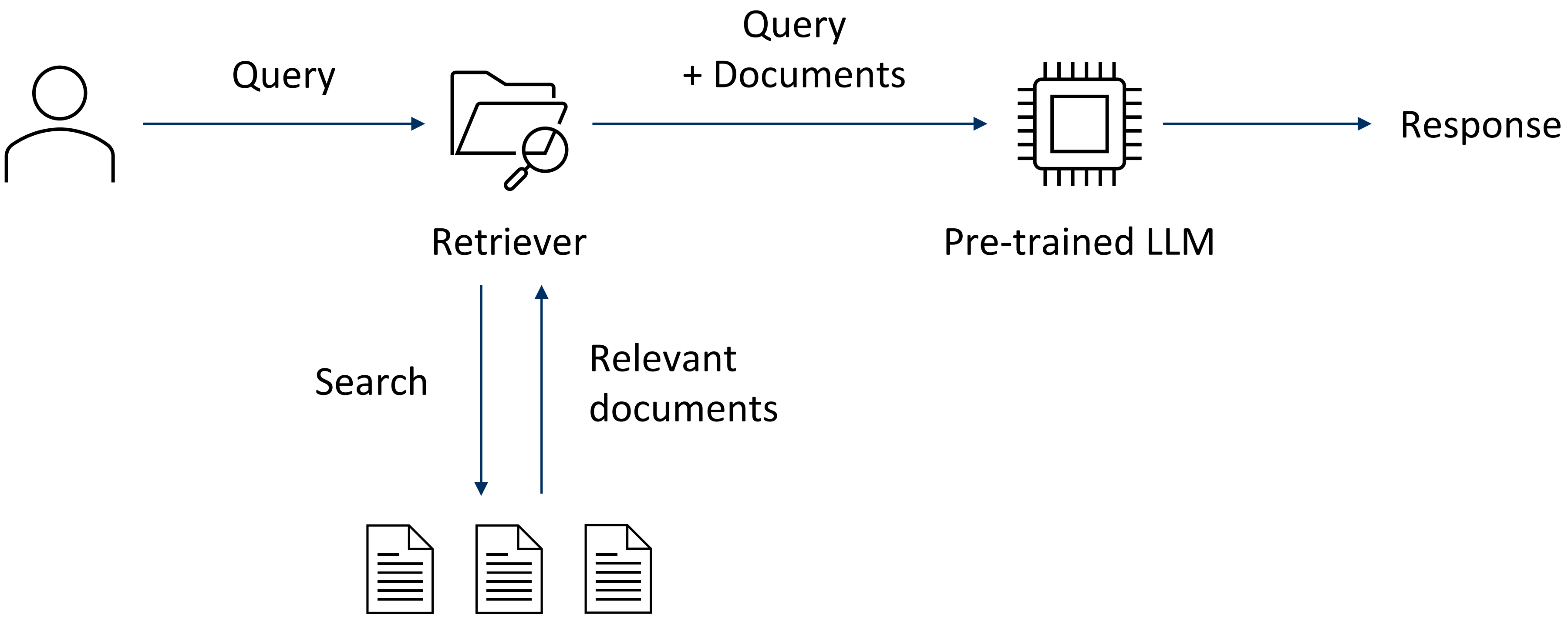




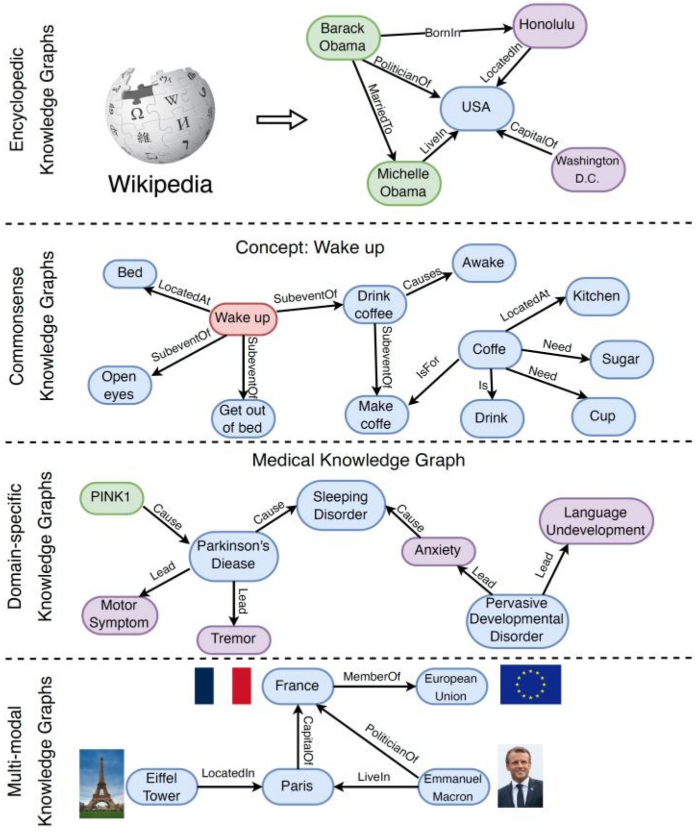
[Prithivi Da, via LinkedIn]

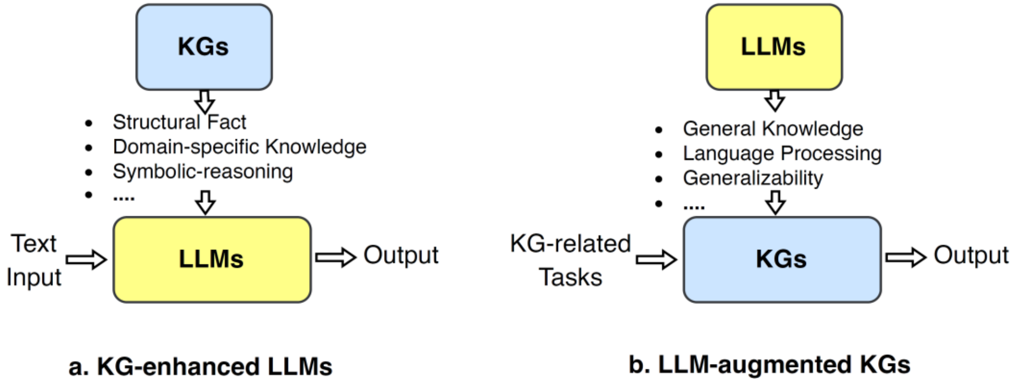
Goal: Improve factual correctness of large language models

Previous Approach: Retrieval Augmented Generation
[Lewis et al. 2020]



Language Models and Knowledge Graphs



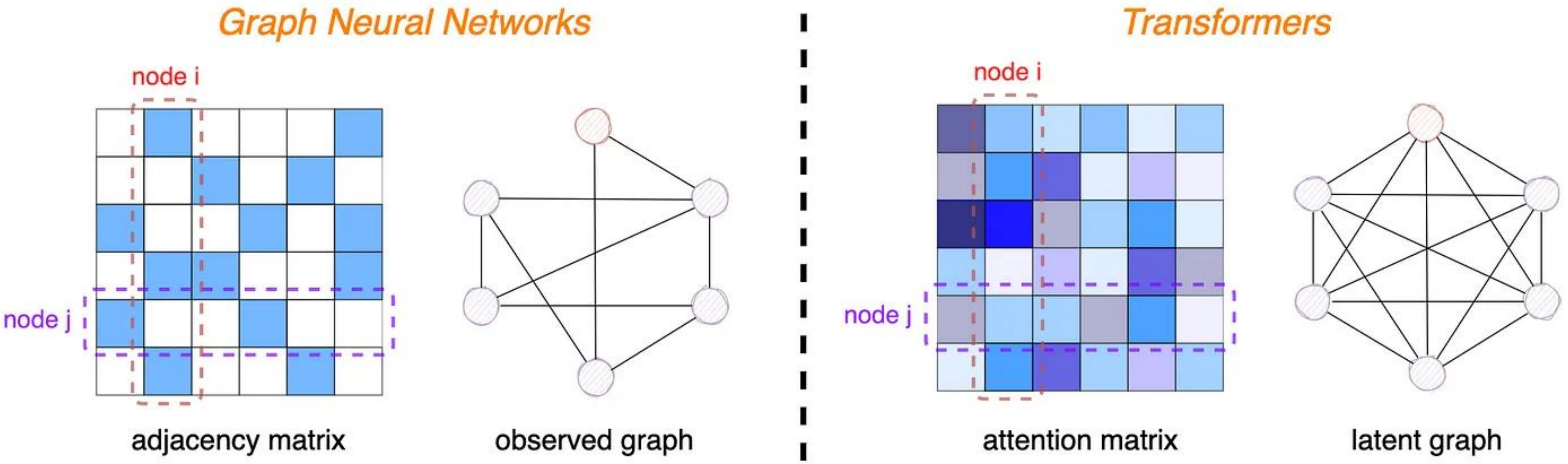


[Pan et al. 2023]

Research question: How to efficiently perform joint inference on text data and symbolic data, e.g. given as a knowledge graph?

Transformer Models as Graph Neural Networks

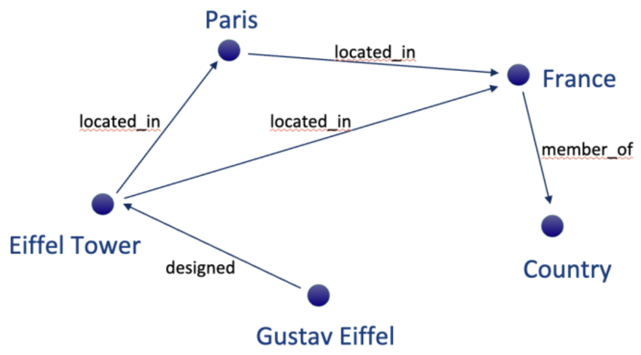
- We take a unifying perspective of Transformer models as Graph Neural Networks on the complete graph:



[Q. Wu 2023]

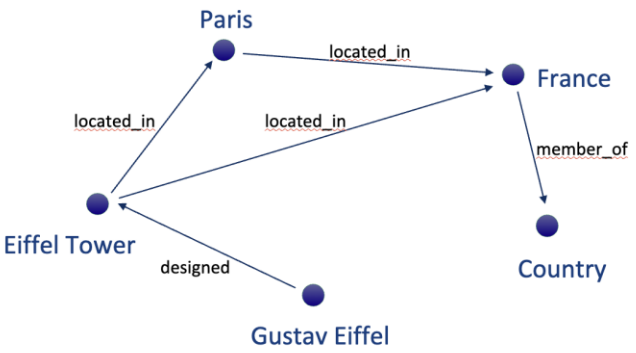
Knowledge Graph enhanced LLMs

Question: In which country is the Eiffel Tower?



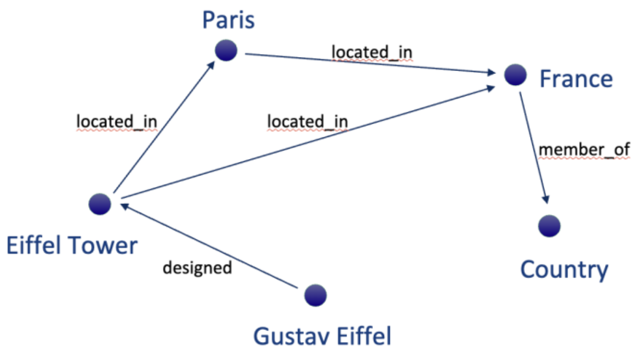
Knowledge Graph enhanced LLMs

Question: In which country
is the Eiffel Tower?



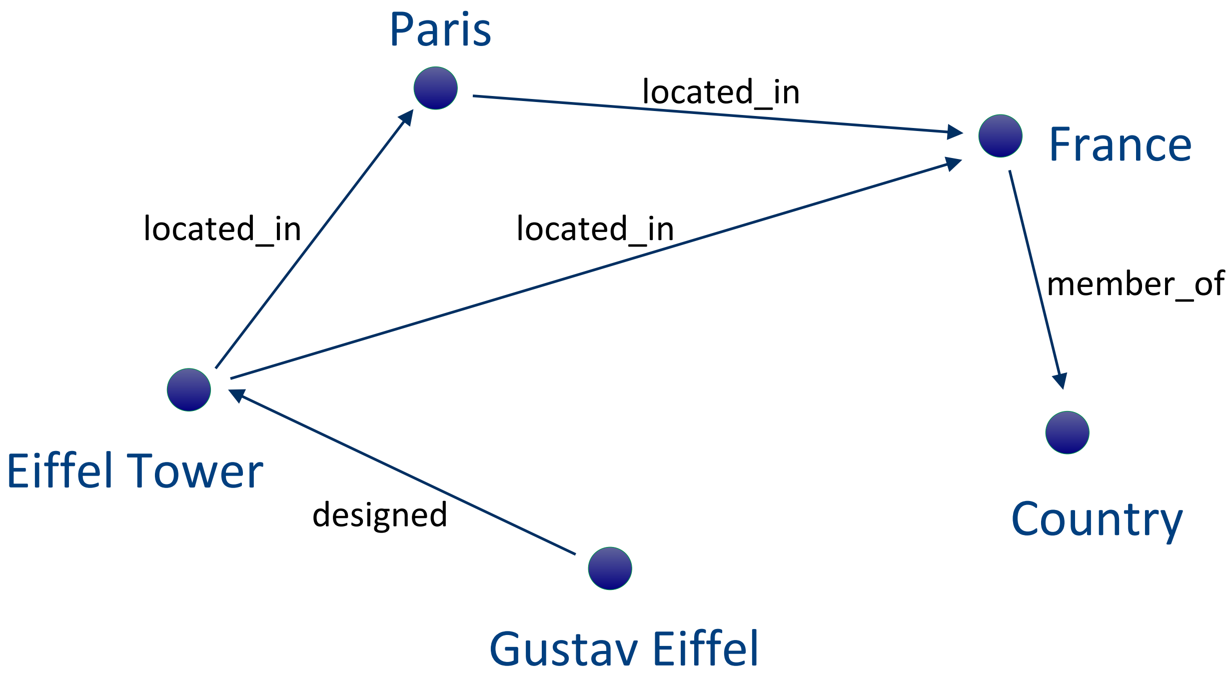
Knowledge Graph enhanced LLMs

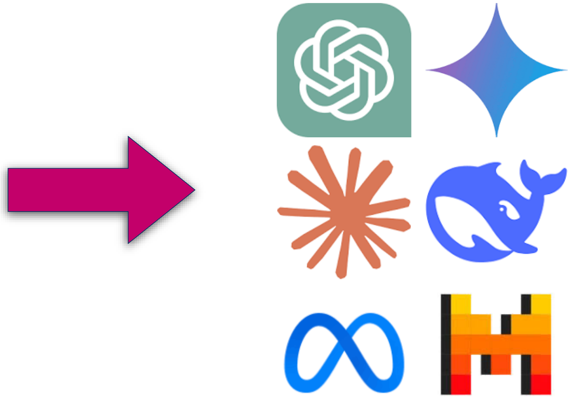
Question: In which country is the Eiffel Tower?



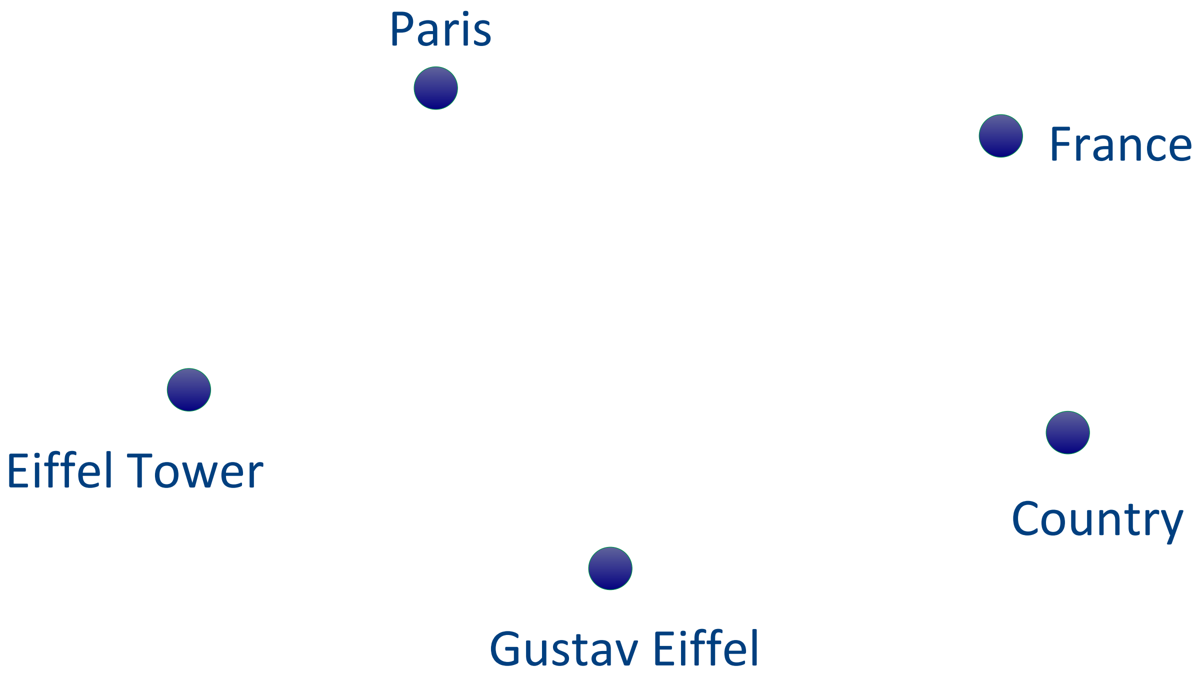
Answer: It is located in France.

How to Input a Graph?

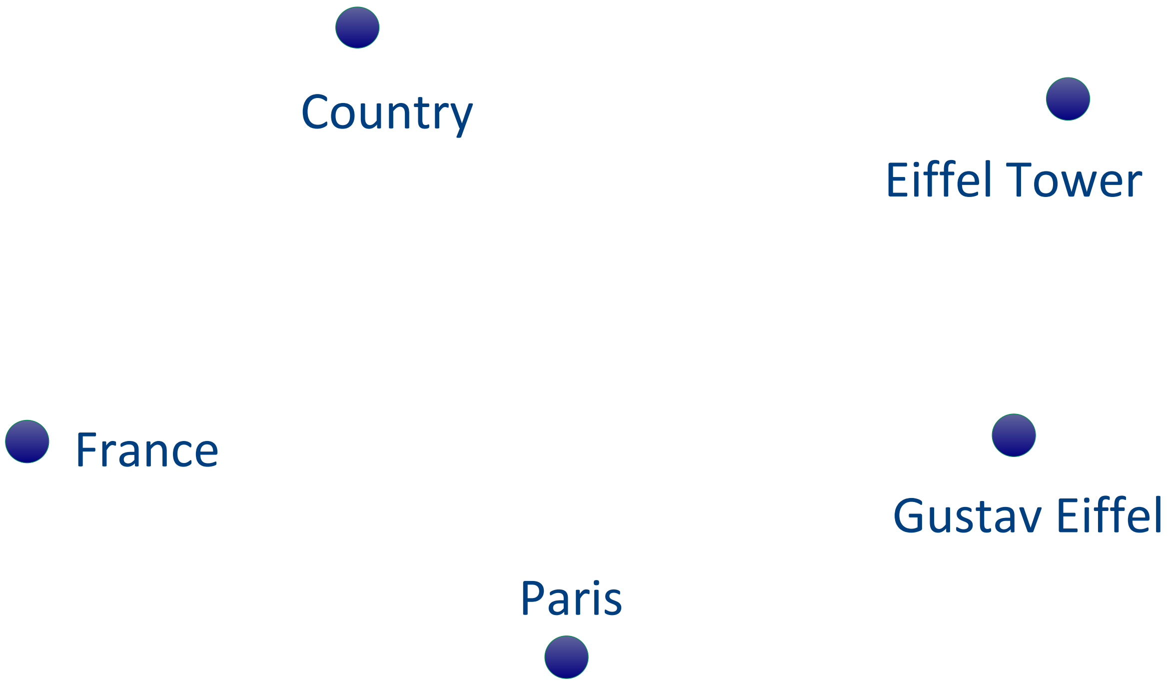




How to Input a ~~Graph~~ Set?

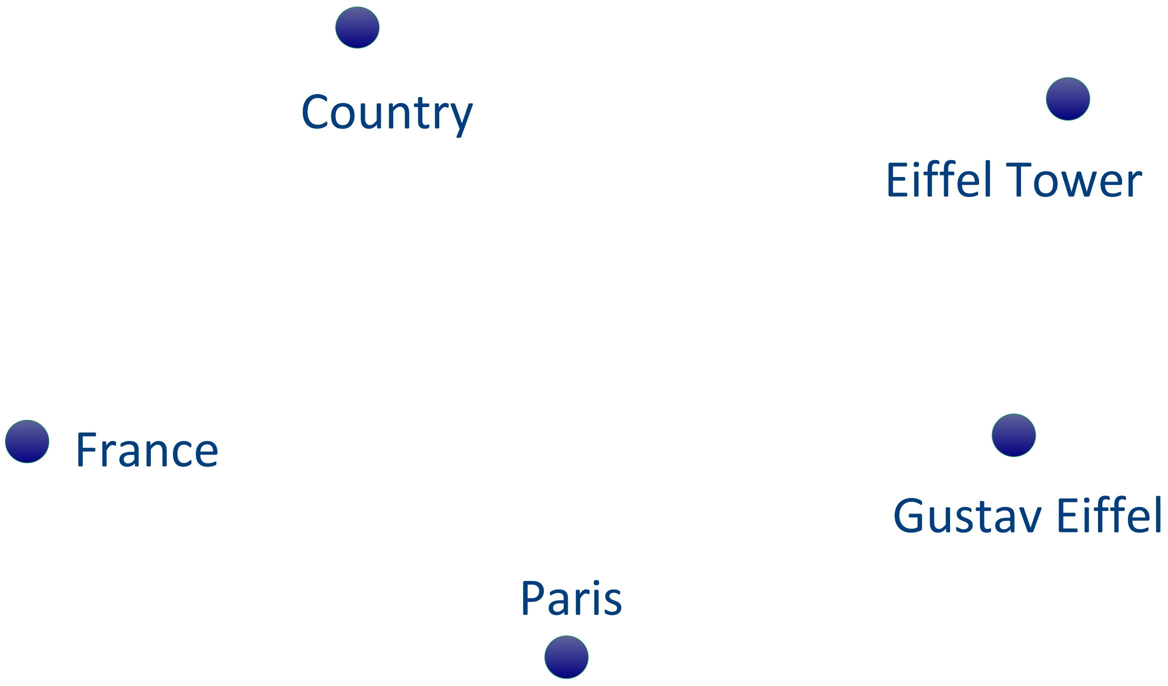


How to Input a ~~Graph~~ Set?



How to Input a ~~Graph~~ Set?

Invariance



Why is Invariance Important?

Question: Where is the Eiffel Tower located?

- Paris
- Heidelberg
- Budapest
- Zurich

Answer: Paris

Why is Invariance Important?

Question: Where is the Eiffel Tower located?

- Heidelberg
- Zurich
- Paris
- Budapest

Answer: Heidelberg

Why is Invariance Important?

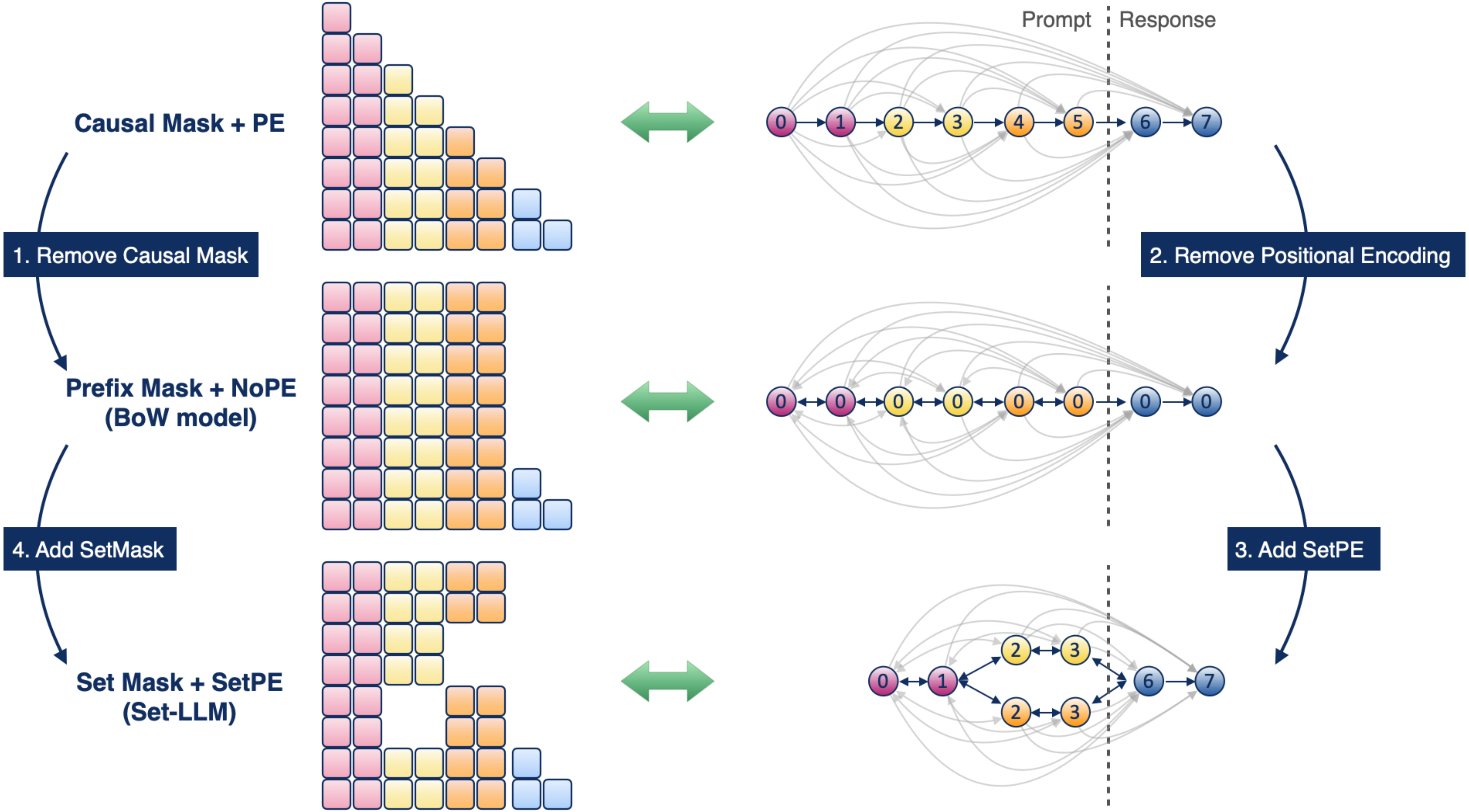
Method	MMLU	ARC-c	BoolQ	SocialiQA	MedMCQA
Llama2-7B	40.91/ 6.17 (34.74 ↓)	47.04/ 7.98 (39.06 ↓)	61.79/ 8.23 (53.56 ↓)	52.00/15.71 (36.29 ↓)	37.96/ 1.60 (36.36 ↓)
Llama2-13B	52.22/18.33 (33.89 ↓)	61.80/21.63 (40.17 ↓)	67.16/38.29 (28.87 ↓)	61.21/34.14 (27.07 ↓)	39.78/ 7.35 (32.43 ↓)
Llama2-70B	64.68/33.16 (31.52 ↓)	80.00/51.50 (28.50 ↓)	76.39/56.21 (20.18 ↓)	71.60/49.85 (21.75 ↓)	49.61/ 7.35 (32.43 ↓)
Vicuna-v1.5	48.57/18.09 (30.48 ↓)	58.37/23.43 (34.94 ↓)	64.04/29.60 (34.44 ↓)	64.99/38.33 (26.66 ↓)	39.28/ 7.67 (31.61 ↓)
Vicuna-v1.5-13B	54.68/26.27 (28.41 ↓)	69.27/38.80 (30.47 ↓)	68.96/42.14 (26.82 ↓)	66.07/44.42 (21.65 ↓)	41.80/11.90 (29.90 ↓)
WizardLM-13B	48.60/15.87 (32.73 ↓)	58.20/21.12 (37.08 ↓)	67.49/42.11 (25.38 ↓)	63.46/31.78 (31.68 ↓)	34.87/ 6.32 (28.55 ↓)
InternLM-7B	45.72/10.45 (35.27 ↓)	56.14/17.34 (38.80 ↓)	65.83/26.41 (39.42 ↓)	59.47/30.30 (29.17 ↓)	32.63/ 2.56 (30.07 ↓)
InternLM-20B	59.14/29.52 (29.62 ↓)	78.28/54.42 (23.86 ↓)	85.20/82.91 (2.29 ↓)	79.48/65.97 (13.51 ↓)	43.61/13.92 (29.69 ↓)
Falcon-7b	31.66/ 2.49 (29.17 ↓)	34.74/ 0.09 (34.65 ↓)	55.35/ 2.66 (52.69 ↓)	36.29/ 0.55 (35.74 ↓)	28.12/ 0.07 (28.05 ↓)
MPT-7B	35.60/ 3.52 (32.08 ↓)	37.76/ 1.06 (36.70 ↓)	58.46/ 7.03 (51.43 ↓)	41.61/ 2.53 (39.08 ↓)	26.31/ 1.60 (24.71 ↓)
GPT-3.5-turbo	64.81/40.39 (24.42 ↓)	82.23/61.55 (20.68 ↓)	87.92/81.35 (6.57 ↓)	70.62/56.29 (14.33 ↓)	52.22/32.07 (20.15 ↓)
Random Chance	25.0	25.0	50.0	33.33	25.0

[1] Zong, Yongshuo, et al. "Fool Your (Vision and) Language Model with Embarrassingly Simple Permutations." (ICML 2024)

Why is Invariance Important?

Method	MMLU	ARC-c	BoolQ	SocialiQA	MedMCQA
Llama2-7B	40.91/ 6.17 (34.74 ↓)	47.04/ 7.98 (39.06 ↓)	61.79/ 8.23 (53.56 ↓)	52.00/15.71 (36.29 ↓)	37.96/ 1.60 (36.36 ↓)
Llama2-13B	52.22/18.33 (33.89 ↓)	61.80/21.63 (40.17 ↓)	67.16/38.29 (28.87 ↓)	61.21/34.14 (27.07 ↓)	39.78/ 7.35 (32.43 ↓)
Llama2-70B	64.68/33.16 (31.52 ↓)	80.00/51.50 (28.50 ↓)	76.39/56.21 (20.18 ↓)	71.60/49.85 (21.75 ↓)	49.61/ 7.35 (32.43 ↓)
Vicuna-v1.5	48.57/18.09 (30.48 ↓)	58.37/23.43 (34.94 ↓)	64.04/29.60 (34.44 ↓)	64.99/38.33 (26.66 ↓)	39.28/ 7.67 (31.61 ↓)
Vicuna-v1.5-13B	54.68/26.27 (28.41 ↓)	69.27/38.80 (30.47 ↓)	68.96/42.14 (26.82 ↓)	66.07/44.42 (21.65 ↓)	41.80/11.90 (29.90 ↓)
WizardLM-13B	48.60/15.87 (32.73 ↓)	58.20/21.12 (37.08 ↓)	67.49/42.11 (25.38 ↓)	63.46/31.78 (31.68 ↓)	34.87/ 6.32 (28.55 ↓)
InternLM-7B	45.72/10.45 (35.27 ↓)	56.14/17.34 (38.80 ↓)	65.83/26.41 (39.42 ↓)	59.47/30.30 (29.17 ↓)	32.63/ 2.56 (30.07 ↓)
InternLM-20B	59.14/29.52 (29.62 ↓)	78.28/54.42 (23.86 ↓)	85.20/82.91 (2.29 ↓)	79.48/65.97 (13.51 ↓)	43.61/13.92 (29.69 ↓)
Falcon-7b	31.66/ 2.49 (29.17 ↓)	34.74/ 0.09 (34.65 ↓)	55.35/ 2.66 (52.69 ↓)	36.29/ 0.55 (35.74 ↓)	28.12/ 0.07 (28.05 ↓)
MPT-7B	35.60/ 3.52 (32.08 ↓)	37.76/ 1.06 (36.70 ↓)	58.46/ 7.03 (51.43 ↓)	41.61/ 2.53 (39.08 ↓)	26.31/ 1.60 (24.71 ↓)
GPT-3.5-turbo	64.81/40.39 (24.42 ↓)	82.23/61.55 (20.68 ↓)	87.92/81.35 (6.57 ↓)	70.62/56.29 (14.33 ↓)	52.22/32.07 (20.15 ↓)
Random Chance	25.0	25.0	50.0	33.33	25.0

[1] Zong, Yongshuo, et al. "Fool Your (Vision and) Language Model with Embarrassingly Simple Permutations." (ICML 2024)



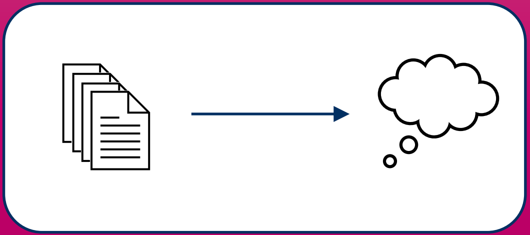
Causal Mask + PE
1. Remove Causal Mask
Prefix Mask + NoPE
(BoW model)
4. Add SetMask
Set Mask + SetPE
(Set-LLM)
Prompt
Response
2. Remove Positional Encoding
3. Add SetPE

Results – Multiple Choice

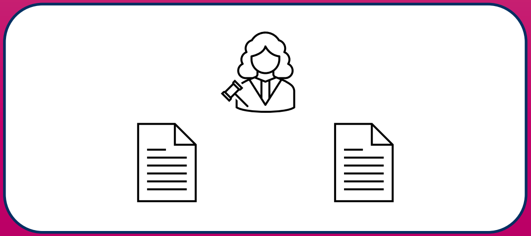
Model	Eval. Mode	# Runs	PIQA		ARC		CSQA		SIQA	
			Rand.	Adv.	Rand.	Adv.	Rand.	Adv.[†]	Rand.	Adv.
Causal Mask+PE	Single run	1	84.11	76.77	55.20	23.72	78.31	69.62	74.80	63.00
Prefix Mask+SetPE	Single run	1	81.23	81.23	51.28	51.28	77.31	77.31	71.24	71.24
SetMask+SetPE	Single run	1	84.33	84.33	57.76	57.76	79.93	79.93	75.38	75.38
SetMask+SetPE$^{\text{Ultra}}$	Single run	1	**85.80**	**85.80**	**65.02**	**65.02**	**80.18**	**80.18**	**76.15**	**76.15**

Applications of Set-LLM beyond Multiple-Choice

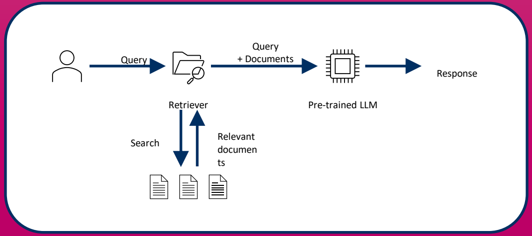
Multi-Document Summarization



LLM as a Judge

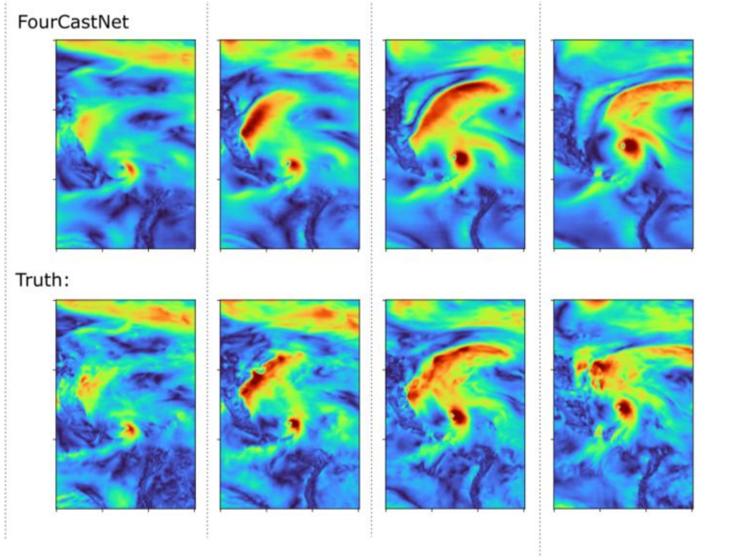


Multi-document Question Answering



PhD opportunity:
Statistical Calibration of Weather and Climate Models

- Improve the statistical calibration of weather and climate models especially for long-range predictions
- Derive statistical calibration of generative models based on proper scoring rules
- Work with multimodal data that combines atmospheric data with e.g. hydrology



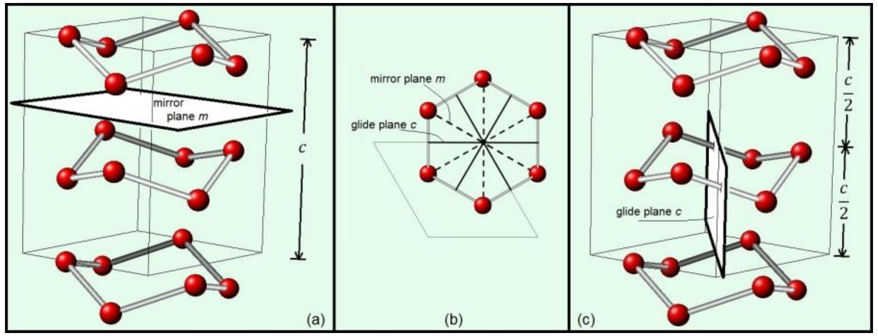
Further Applications:

- Extends to generative models for e.g. proteins

PhD opportunity:
Equivariant Models for Materials Science

- Include space group symmetries in generative models for crystalline materials
- End-to-end learning of a multi-resolution hierarchy for molecular property prediction

Important Applications:
- Design of catalysts for direct carbon air capture



[Wikipedia, Dbuckingham42, CC BY-SA 4.0]

T. Berndt, J. Stühmer. “Approximate equivariance via projection-based regularization”, ICML 2026

B. Egressy, J. Stühmer. “Set-LLM: A permutation-invariant LLM”. NeurIPS 2025

S. Wagner*, L. Seute*, V. Viliuga, N. Wolf, F. Gräter, J. Stühmer. “Generating highly designable proteins with geometric algebra flow matching”. NeurIPS 2024

V. Viliuga*, L. Seute*, N. Wolf, S. Wagner, A. Elofsson, J. Stühmer, F. Gräter. “Flexibility-conditioned protein structure design with flow matching”. ICML 2025.

N. Wolf*, L. Seute*, V. Viliuga, S. Wagner, J. Stühmer, F. Gräter. “Learning conformational ensembles of proteins based on backbone geometry”. NeurIPS 2025