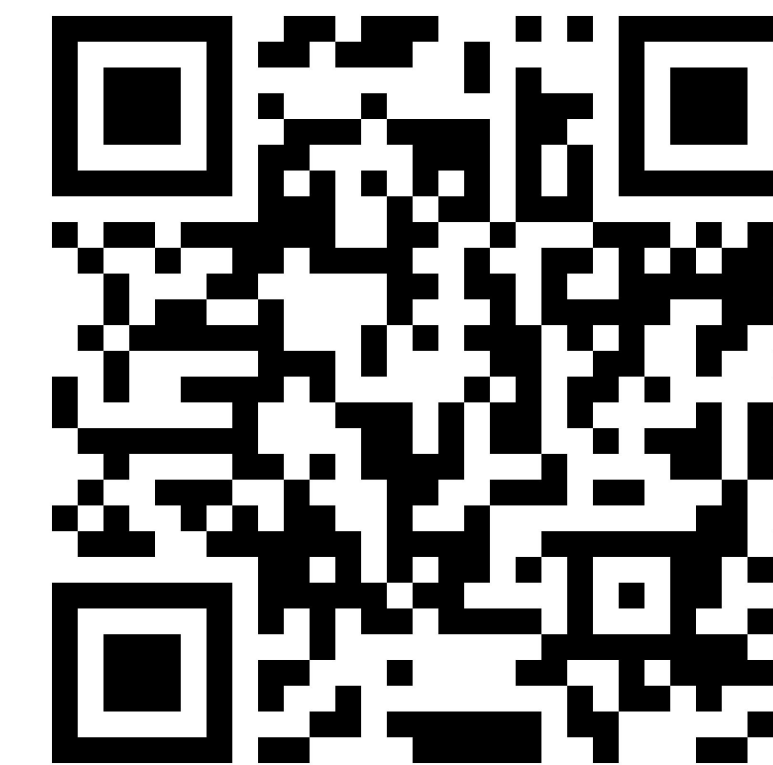
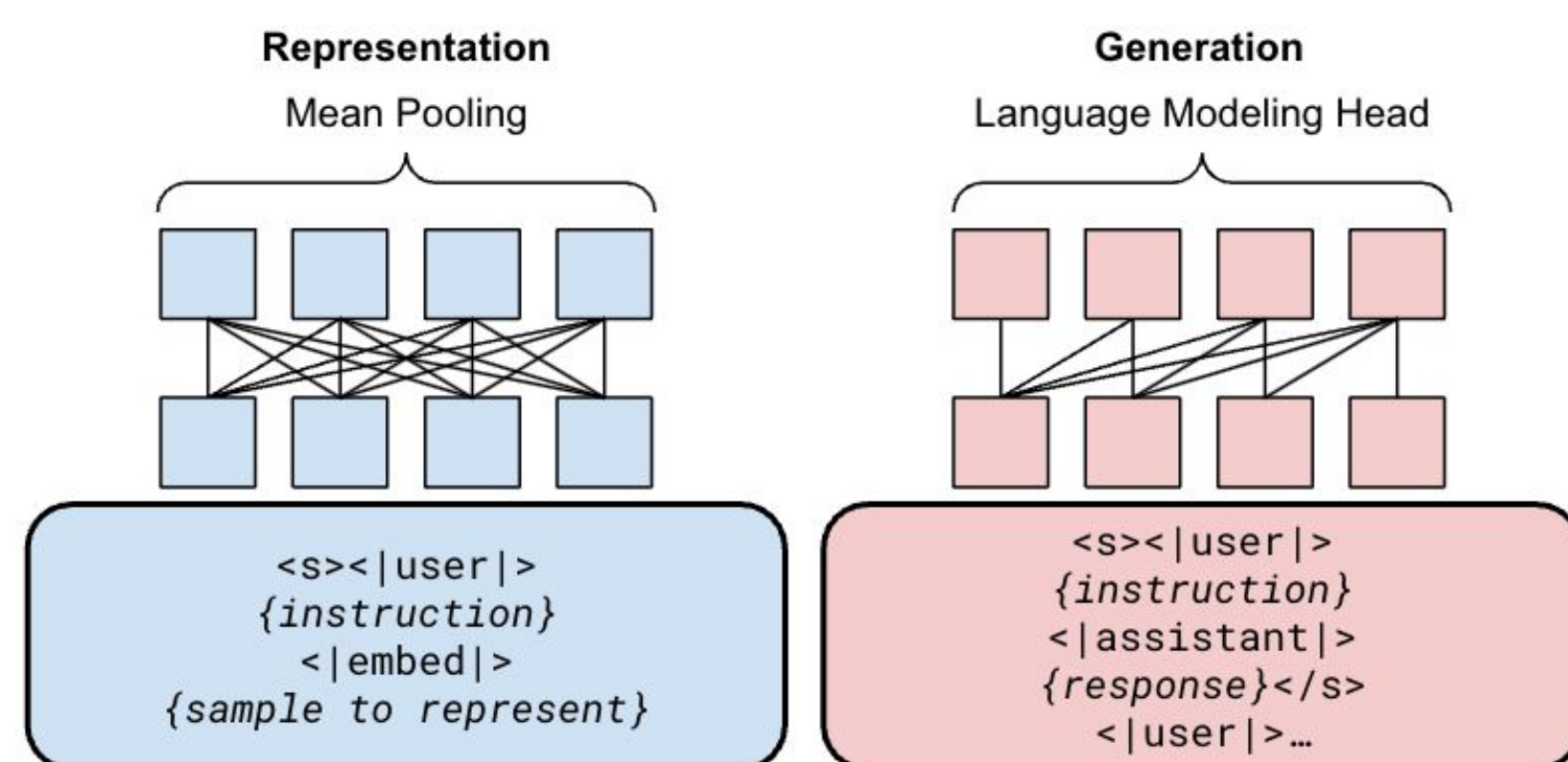
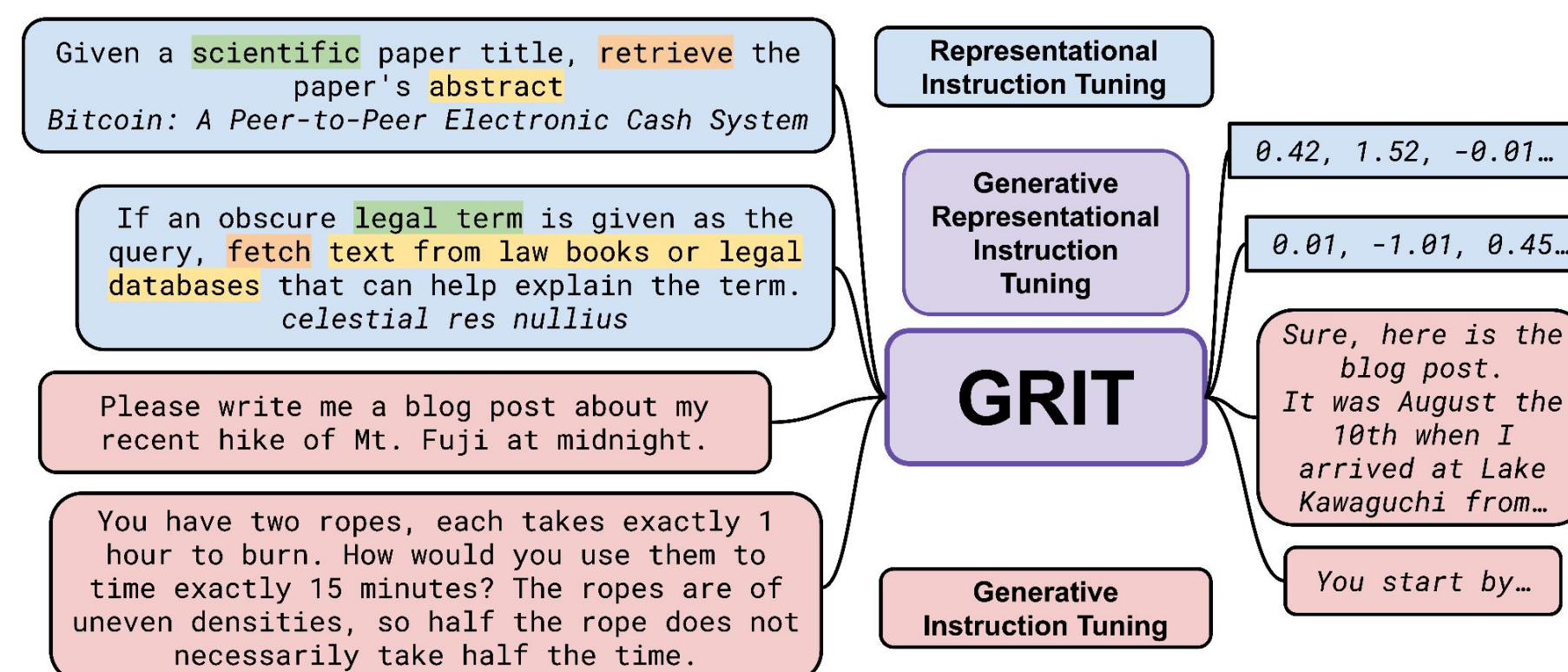
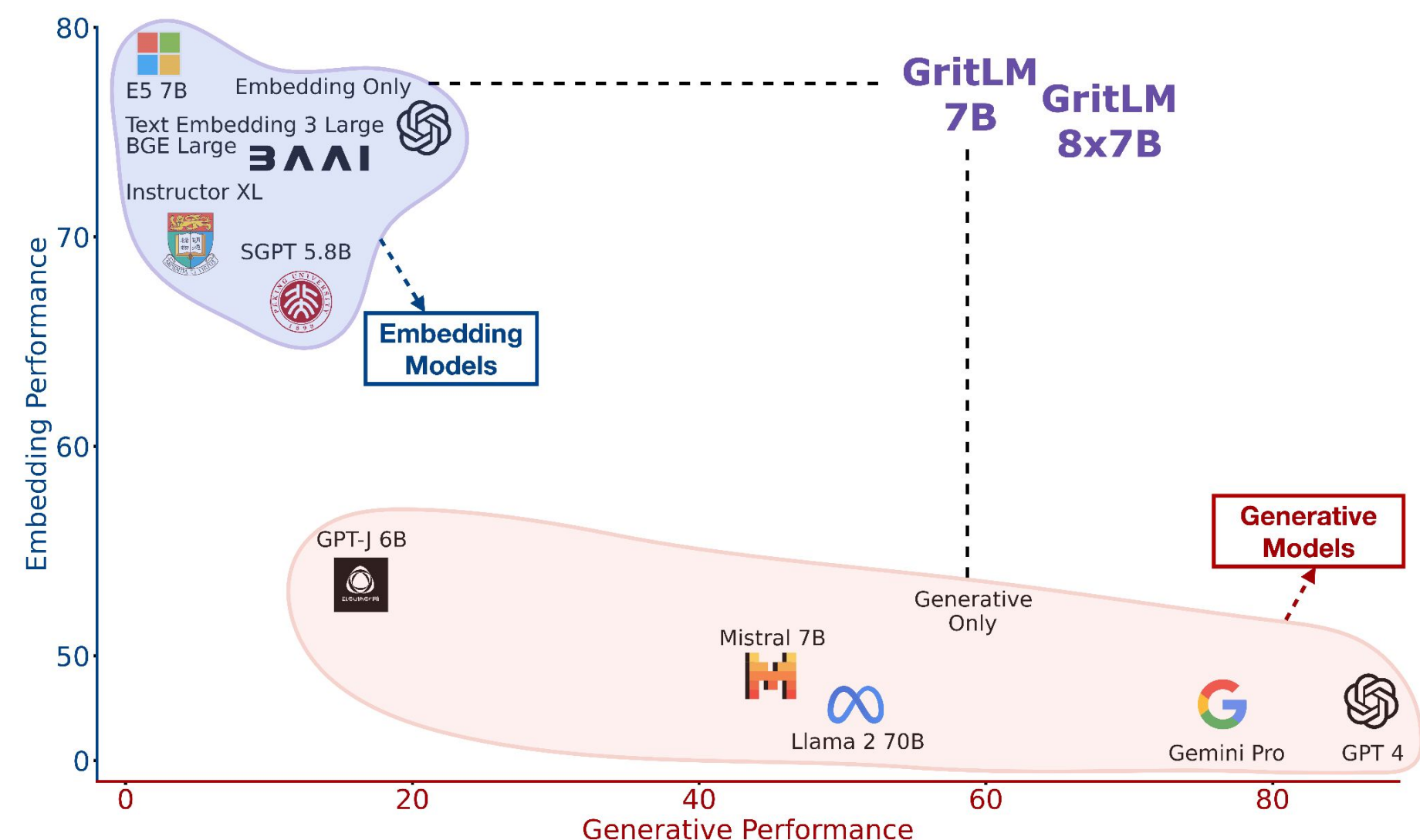


# Generative Representational Instruction Tuning

Niklas Muennighoff, Hongjin Su, Liang Wang, Nan Yang, Furu Wei, Tao Yu, Amanpreet Singh, Douwe Kiela



## GRIT: Generative Representational Instruction Tuning



**objective**

$$\mathcal{L}_{\text{Rep}} = -\frac{1}{M} \sum_{i=1}^M \log \frac{\exp(\tau \cdot \sigma(f_{\theta}(q^{(i)}), f_{\theta}(d^{(i)})))}{\sum_{j=1}^M \exp(\tau \cdot \sigma(f_{\theta}(q^{(i)}), f_{\theta}(d^{(j)})))}$$
$$\mathcal{L}_{\text{Gen}} = -\frac{1}{N} \sum_{i=1}^N \log P(f_{\theta, \eta}(x^{(i)}) | f_{\theta, \eta}(x^{(<i>{</i>})})$$
$$\mathcal{L}_{\text{GRIT}} = \lambda_{\text{Rep}} \mathcal{L}_{\text{Rep}} + \lambda_{\text{Gen}} \mathcal{L}_{\text{Gen}}$$

## GritLM on embedding and generation

### Embedding

Task (→)	CLF	Clust.	PairCLF	Rerank	Retrieval	STS	Summ.	Avg.
Metric (→)	Acc.	V-Meas.	AP	MAP	nDCG	Spear.	Spear.	
Dataset # (→)	12	11	3	4	15	10	1	56
Proprietary models♥								
OpenAI v3	75.5	49.0	85.7	59.2	55.4	81.7	29.9	64.6
Other Open Models♥								
Llama 2 70B	60.4	29.0	47.1	38.5	9.0	49.1	26.1	35.6
Mistral 7B	63.5	34.6	53.5	43.2	13.2	57.4	19.7	40.5
Mistral 7B Instruct	67.1	34.6	59.6	44.8	16.3	63.4	25.9	43.7
GPT-J 6B	66.2	39.0	60.6	48.9	19.8	60.9	26.3	45.2
SGPT BE 5.8B	68.1	40.3	82.0	56.6	50.3	78.1	31.5	58.9
Instructor XL 1.5B	73.1	44.7	86.6	57.3	49.3	83.1	<b>32.3</b>	61.8
BGE Large 0.34B	76.0	46.1	87.1	60.0	54.3	83.1	31.6	64.2
E5 Mistral 7B	78.5	50.3	<b>88.3</b>	60.2	56.9	<b>84.6</b>	31.4	66.6
GRITLM								
Gen.-only 7B	65.4	32.7	54.2	43.0	13.7	60.2	21.1	41.2
Emb.-only 7B	78.8	<b>51.1</b>	87.1	<b>60.7</b>	<b>57.5</b>	83.8	30.2	<b>66.8</b>
GRITLM 7B	<b>79.5</b>	<u>50.6</u>	<u>87.2</u>	<u>60.5</u>	<u>57.4</u>	83.4	30.4	<b>66.8</b>
GRITLM 8x7B	78.5	50.1	85.0	59.8	55.1	83.3	29.8	65.7

### Generation

Dataset (→)	MMLU	GSM8K	BBH	TyDi QA	HumanEval	Alpaca	Avg.
Setup (→)	0 FS	8 FS, CoT	3 FS, CoT	1 FS, GP	0 FS	0 FS, 1.0	
Metric (→)	EM	EM	EM	F1	pass@1	% Win	
Proprietary models♥							
GPT-4-0613	81.4	95.0	89.1	65.2	86.6 <sup>†</sup>	91.2	84.8
Other Open Models♥							
GPT-J 6B	27.7	2.5	30.2	9.4	9.8	0.0	13.3
SGPT BE 5.8B	24.4	1.0	0.0	22.8	0.0	0.0	8.0
Zephyr 7B β	58.6	28.0	44.9	23.7	28.5	85.8	44.9
Llama 2 7B	41.8	12.0	39.3	51.2	12.8♦	0.0	26.2
Llama 2 13B	52.0	25.0	48.9	56.5	18.3♦	0.0	33.5
Llama 2 70B	64.5	55.5	66.0	<b>62.6</b>	29.9♦	0.0	46.4
Llama 2 Chat 13B	53.2	9.0	40.3	32.1	19.6 <sup>†</sup>	91.4	40.9
Llama 2 Chat 70B	60.9	59.0	49.0	44.4	34.3 <sup>†</sup>	<u>94.5</u>	57.0
Tulu 2 7B	50.4	34.0	48.5	46.4	24.5 <sup>†</sup>	73.9	46.3
Tulu 2 13B	55.4	46.0	49.5	53.2	31.4	78.9	52.4
Tulu 2 70B	<u>67.3</u>	<b>73.0</b>	68.4	53.6	41.6	86.6	<u>65.1</u>
Mistral 7B	60.1	44.5	55.6	55.8	30.5	0.0	41.1
Mistral 7B Instruct	53.0	36.0	38.5	27.8	34.0	75.3	44.1
Mixtral 8x7B Instruct	<b>68.4</b>	<u>65.0</u>	55.9	24.3	<b>53.5</b>	<b>94.8</b>	60.3
GRITLM							
Emb.-only 7B	23.5	1.0	0.0	21.0	0.0	0.0	7.6
Gen.-only 7B	57.5	52.0	55.4	56.6	34.5	75.4	55.2
GRITLM 7B	57.6	57.5	54.8	55.4	32.8	74.8	55.5
GRITLM 8x7B	66.7	61.5	<b>70.2</b>	58.2	<u>53.4</u>	84.0	<b>65.7</b>

Best-in-class performance in a single unified model at both types of tasks simultaneously

## RAG with GRIT

