**Socialite-Llama: An instruction-Tuned Lamma2 for Social Scientific Tasks**

**Gourab Dey and Adithya Ganesan**

**Introduction.** We will introduce Socialite- Llama, a Llama2 7B-based LLM (Touvron et al., 2023) instruction tuned on a suite of social scientific tasks for which we carefully handcraft instructions. Social science-related NLP tasks, such as emotion or humor detection, capture important language semantics to add to the implicit pragmatics from the text. Instruction- finetuned large language models (LLMs) have demonstrated impressive performance on many standard NLP tasks (Wei et al., 2022; Chung et al., 2022), but these models, tuned on non-social tasks, seem to have poor social pragmatics (Ziems et al., 2023; Choi et al., 2023).

**Background.** Instruction tuning has been successful with limited amounts of data and instruction-tuned models can generalize to new tasks (Gupta et al., 2023). This suggests that LLMs could become more socially capable by instruction tuning them on a wide variety of social NLP tasks.

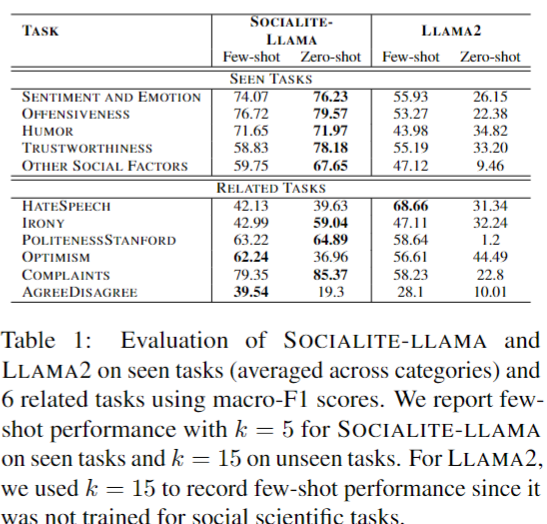
**Dataset.**

* 26 total tasks, spanning > 100m tokens; > 50GB, taken from the SOCKET evaluation (Choi et al., 2023). *(note to NLP class students: this is probably more than you should try)*
* Train set:: 20 different task datasets across 5 broad categories: Humor, Offensiveness, Sentiment and Emotion, Trustworthiness, and others.
* Test Set: our model’s performance on the 20 seen tasks as well as 6 unseen but related social tasks.

**Method.** We will develop instructions for each of the 20 training tasks and the 6 related social tasks and send them through tokenization **[1. Syntax]**.Instructions will be designed to utilize dependency parsing to identify the answer to questions **[2. Semantics].** We will train the LM **[3. Transformer LLMs]** on the 20 training tasks with the instructions and training set data. We will then test on held-out test tests of the training data as well as the 6 unseen but similar tasks modeling psychology and behavior **[4. Applications]**. Both of us will design the instruction set. Adi will design the approach and train the LLMs while Gourab will carry out the tests and evaluations.

**Potential Results.** We will compare Socialite to Lamma2 and DeBerta, and attempt to show that it significantly outperforms other open models. We will compare performance seen tasks and the unseen related tasks.

**Take aways.** We hope to show that by instruction-tuning LLMs for social tasks they can learn to be stronger at additional social tasks, even those unseen during training.





**References.**

Minje Choi, Jiaxin Pei, Sagar Kumar, Chang Shu, and David Jurgens. 2023. Do LLMs understand social knowledge? evaluating the sociability of large language models with SocKET benchmark. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing,* pp. 11370–11403.

Hyung Won Chung, Le Hou, …, and Jason Wei. 2022. Scaling instruction-finetuned language models. *Journal of Machine Learning Research.*

Himanshu Gupta, Saurabh Arjun Sawant, Swaroop Mishra, Mutsumi Nakamura, Arindam Mitra, Santosh Mashetty, and Chitta Baral. 2023. Instruction tuned models are quick learners. *arXiv preprint arXiv:2306.05539.*

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288.*

Jason Wei, Maarten Bosma, Vincent Y. Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M. Dai, and Quoc V. Le. 2022. Finetuned language models are zero-shot learners. *International Conference on Learning Representations.*

Caleb Ziems, William Held, Omar Shaikh, Jiaao Chen, Zhehao Zhang, and Diyi Yang. 2023. Can Large Language Models Transform Computational Social Science? *Computational Linstuistics. pp 1 - 53.*