Foundation Models

Technical Advances, Social Responsibility, Applications

PIs: Percy Liang, Russ Altman Jeannette Bohg, Akshay Chauhari, Chelsea Finn, Tatsunori Hashimoto, Daniel Ho, Fei-Fei Li, Chris Manning, Tengyu Ma, Chris Ré, Rob Reich, Dorsa Sadigh, Matei Zaharia

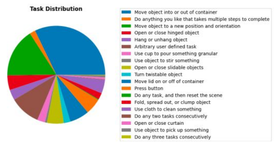
Center for Research on Foundation Models

Applications

[Sasha Khazatsky, …, **Jeannette Bohg, Dorsa Sadigh, Chelsea Finn**]

Household robotics dataset

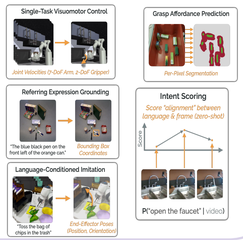
- Goal: data from many distinct environments
- 70 people from 18 organizations
- 50+ real households with 20 scenes each
- 30K demonstrations (goal: 100K)



[Siddharth Karamcheti, Suraj Nair, Annie S. Chen, Thomas Kollar, **Chelsea Finn**, **Dorsa Sadigh**, **Percy Liang**]

Voltron (robotic foundation model)

- Trained on Something-Something v2 (220K human video clips)
- Use text conditioning and generation to capture low-level spatial features and high-level semantics
- Evaluate on 5 diverse robotic tasks:



[Pierre Chambon, Christian Bluethgen, Jean-Benoit Delbrouck, …, Curtis P. Langlotz, **Akshay Chaudhari**]

RoentGen (text-to-image model)

- Generates synthetic chest x-rays given natural language description
- Fine-tune Stable Diffusion on medical images
- Uses: data augmentation for improving classifiers, new teaching tool

[Neel Guha . . . Julian Nyarko, **Daniel E. Ho**, **Christopher Ré**]

LegalBench (evaluation)

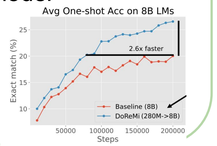
- Legal reasoning: issue spotting, rule recall, analysis, conclusion
- **162 tasks** (and growing!) contributed by ~40 legal experts across civil justice areas
- Evaluation on 20 FMs from 11 different families of models

Technical Advances

[Sang Michael Xie, Hieu Pham, Xuanyi Dong, Nan Du, Hanxiao Liu, Yifeng Lu, **Percy Liang**, Quoc V. Le, **Tengyu Ma**, Adams Wei Yu]

Domain Reweighting with Minimax Optimization (DoReMi)

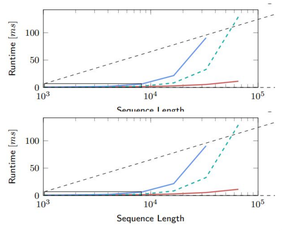
- Uses distributionally robust optimization (DRO) to optimize domain weights on a small 280M proxy model
- Use domain weights to train a 8B model
- **2.6x** faster (cheaper) than using heuristic domain weights

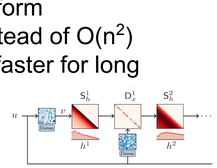


[Michael Poli, Stefano Massaroli, Eric Nguyen, Daniel Y. Fu, Tri Dao, Stephen Baccus, Yoshua Bengio, Stefano Ermon, **Christopher Ré**]

Hyena

- Long convolutions using Fast Fourier Transform
- O(n log n) instead of $O(n^2)$
- Result: **100x** faster for long sequences

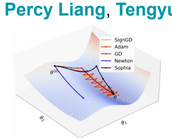


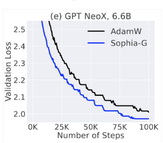


[Hong Liu, Zhiyuan Li, David Hall, **Percy Liang**, **Tengyu Ma**]

Sophia: Second-order Clipped Stochastic Optimization

- Diagonal approximation of the Hessian (second-order) with clipping
- Outperforms Adam by 2x for LLMs from scratch on models from 125M to 6.7B parameters

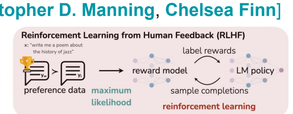


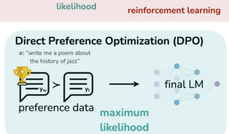


[Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, **Christopher D. Manning**, **Chelsea Finn**]

Direct Policy Optimization (DPO)

- Alignment is an important final step for endowing LMs with instruction following and safety
- Simpler (no reward model, no sampling completions, no hyperparameter tuning)
- Perform at least as well as RLHF





Social Responsibility

[**Percy Liang**, Rishi Bommasani, Tony Lee, …]

Holistic Evaluation of Language Models (HELM)

- A **reproducible** and **transparent** framework for evaluating foundation models.
- Leaderboards with **many scenarios, metrics, and models**

HELM Lite →
Lightweight, broad evaluation of the capabilities of language models using in-context learning

HELM Classic →
Thorough language model evaluations based on the scenarios from the original HELM paper

HEIM →
Holistic evaluation of text-to-image models

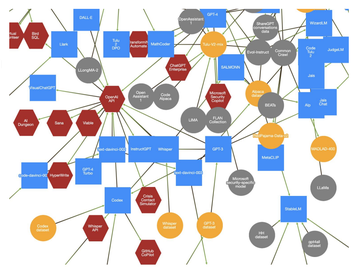
HELM Instruct →
Evaluations of instruction following models with absolute ratings

HELM MMLU →
Massive Multitask Language Understanding (MMLU) evaluations using standardized prompts

VHELM →
Holistic Evaluation of Vision-Language Models

FM Ecosystem Graphs

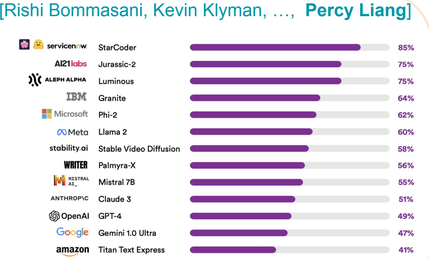
- Foundation models encompass an ecosystem of datasets, models, and applications.
- This framework documents assets (datasets, models, applications) and their relationships within the foundation models ecosystem.
- Aids researchers, developers, policymakers, and the public in tracking foundation models, their creators, usage trends, and overall ecosystem dynamics.



Foundation Model Transparency Index (FMTI)

[Rishi Bommasani, Kevin Klyman, …, **Percy Liang**]

- Measures transparency of major developers on 100 indicators.
- Originally, developers scored an average of 37/100, but a follow-up showed improved scores: 14 developers averaged 58/100.
- Systemic opacity persists in areas like copyright status and data access.



Acknowledgments: We gratefully acknowledge the support of the Hoffmann-Yee foundation for this work.