

: Accelerating the Science of Language Models

Dirk Groeneveld Iz Beltagy

Pete Walsh Akshita Bhagia Rodney Kinney Oyvind Tafjord

Ananya Harsh Jha Hamish Ivison^β Ian Magnusson Yizhong Wang^β

Shane Arora David Atkinson Russell Authur Khyathi Raghavi Chandu

Arman Cohan^γ Jennifer Dumas Yanai Elazar^β Yuling Gu

Jack Hessel Tushar Khot William Merrill^δ Jacob Morrison

Niklas Muennighoff Aakanksha Naik Crystal Nam Matthew E. Peters

Valentina Pyatkin^β Abhilasha Ravichander Dustin Schwenk Saurabh Shah

Will Smith Emma Strubell^μ Nishant Subramani Mitchell Wortsman^β

Pradeep Dasigi Nathan Lambert Kyle Richardson

Luke Zettlemoyer^β Jesse Dodge Kyle Lo Luca Soldaini

Noah A. Smith^β Hannaneh Hajishirzi^β

Allen Institute for Artificial Intelligence

^βUniversity of Washington ^γYale University

^δNew York University ^μCarnegie Mellon University

olmo@allenai.org

Abstract

Language models (LMs) have become ubiquitous in both NLP research and in commercial product offerings. As their commercial importance has surged, the most powerful models have become closed off, gated behind proprietary interfaces, with important details of their training data, architectures, and development undisclosed. Given the importance of these details in scientifically studying these models, including their biases and potential risks, we believe it is essential for the research community to have access to powerful, truly open LMs. To this end, we have built OLMo, a competitive, truly Open Language Model, to enable the scientific study of language models. Unlike most prior efforts that have only released model weights and inference code, we release OLMo alongside open training data and training and evaluation code. We hope this release will empower the open research community and inspire a new wave of innovation.

1 Introduction

Language models have been at the center of NLP technologies for many years (Rosenfeld, 2000; Ben-

gio et al., 2003; Mikolov et al., 2013; Peters et al., 2018; Brown et al., 2020). Recently, due to large-scale pretraining and human annotation for alignment, they have become commercially valuable (OpenAI, 2023). However, as their commercial value has increased, the largest models have become gated behind proprietary interfaces, with important details left undisclosed.

We believe that full access to open language models for the research community is critical to the scientific study of these models, their strengths and weaknesses, and their biases and risks. Accordingly, we introduce **OLMo**, a powerful, truly open language model alongside open training data, training and evaluation code, intermediate model checkpoints, and training logs.

Recent LM releases have varied in their degree of openness. For example, Mixtral 8x7B provided model weights and a brief report (Jiang et al., 2024), while LLaMA came with in-depth adaptation training instructions (Touvron et al., 2023b), and Mosaic Pretrained Transformer came with many details, including the dataset distribution, though not the data itself (MosaicML NLP Team,