CoNLL 2023

The BabyLM Challenge at the 27th Conference on Computational Natural Language Learning

Proceedings of the BabyLM Challenge

December 6-7, 2023

©2023 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-952148-02-6

Introduction

Greetings "babies"! and welcome to the proceedings and session of the 2023 BabyLM Challenge, held on December 6, 2023 as part of CoNLL (co-hosted with EMNLP) in Singapore. This challenge aims to bring together researchers interested in developmentally plausible pre-training, sample efficiency, and human language acquisition. Our challenge encourages researchers to "think small" by using training corpora containing 100 million words—approximately the amount of data available to human language learners, but far less data than is typically used for pre-training language models.

We received 31 papers, all of which were accepted on the basis of scientific and technical validity, rather than model performance. We received 162 individual model submissions, the scores of which are hosted online, at www.https://dynabench.org/babylm.

We are grateful to the participants for advancing our understanding of how best to train language models on scaled-down and more developmentally plausible corpora.. Their contributions have provided insight into important questions related to cognitive modeling, computational psycholinguistics, and sample-efficient language modeling. We are also grateful to the program committee for their thoughtful reviews of the submissions we received this year. Likewise, we are thankful to the CoNLL organizers for their work in integrating the BabyLM challenge into their program.

– The BabyLM Organization Committee: Alex Warstadt, Aaron Mueller, Leshem Choshen, Ethan Wilcox, Chengxu Zhuang, Adina Williams, Tal Linzen, Ryan Cotterell.

Organizing Committee

Organizers

Alex Warstadt, ETH Zürich, Switzerland
Aaron Mueller, Northeastern University
Leshem Choshen, MIT, IBM
Ethan Wilcox, ETH Zürich, Switzerland
Chengxu Zhuang, Massachusetts Institute of Technology
Juan Ciro, MLCommons
Rafael Mosquera, MLCommons
Bhargavi Paranjabe, University of Washington
Adina Williams, Meta AI (FAIR)
Tal Linzen, New York University
Ryan Cotterell, ETH Zürich, Switzerland

Table of Contents

Findings of the BabyLM Challenge: Sample-Efficient Pretraining on Developmentally Plausible Corpora
Alex Warstadt, Aaron Mueller, Leshem Choshen, Ethan Wilcox, Chengxu Zhuang, Juan Ciro, Rafael Mosquera, Bhargavi Paranjabe, Adina Williams, Tal Linzen and Ryan Cotterell 1

GPT-wee: How Small Can a Small Language Model Really Get?
Bastian Bunzeck and Sina Zarrieß 35

Tiny Language Models Enriched with Multimodal Knowledge from Multiplex Networks
Clayton Fields, Osama Natouf, Andrew McMains, Catherine Henry and Casey Kennington . . . 47

Mini Minds: Exploring Bebeshka and Zlata Baby Models
Irina Proskurina, Guillaume Metzler and Julien Velcin 58

Grammar induction pretraining for language modeling in low resource contexts
Xuanda Chen and Eva Portelance 69

ChapGTP, ILLC's Attempt at Raising a BabyLM: Improving Data Efficiency by Automatic Task Formation
Jaap Jumelet, Michael Hanna, Marianne de Heer Kloots, Anna Langedijk, Charlotte Pouw and Oskar van der Wal 74

Penn & BGU BabyBERTa+ for Strict-Small BabyLM Challenge
Yahan Yang, Elior Sulem, Insup Lee and Dan Roth 86

Too Much Information: Keeping Training Simple for BabyLMs
Lukas Edman and Lisa Bylinina 89

Can training neural language models on a curriculum with developmentally plausible data improve alignment with human reading behavior?
Aryaman Chobey, Oliver Smith, Anzi Wang and Grusha Prasad 98

CLIMB – Curriculum Learning for Infant-inspired Model Building
Richard Diehl Martinez, Hope McGovern, Zebulon Goriely, Christopher Davis, Andrew Caines, Paula Buttery and Lisa Beinborn 112

Acquiring Linguistic Knowledge from Multimodal Input
Theodor Amariucai and Alexander Scott Warstadt 128

Large GPT-like Models are Bad Babies: A Closer Look at the Relationship between Linguistic Competence and Psycholinguistic Measures
Julius Steuer, Marius Mosbach and Dietrich Klakow 142

Baby's CoThought: Leveraging Large Language Models for Enhanced Reasoning in Compact Models
Zheyu Zhang, Han Yang, Bolei Ma, David Rügamer and Ercong Nie 158

ToddlerBERTa: Exploiting BabyBERTa for Grammar Learning and Language Understanding
Ömer Veysel Çağatan 171

CogMemLM: Human-Like Memory Mechanisms Improve Performance and Cognitive Plausibility of LLMs
Lukas Thoma, Ivonne Weyers, Erion Çano, Stefan Schweter, Jutta L Mueller and Benjamin Roth 180

BabyStories: Can Reinforcement Learning Teach Baby Language Models to Write Better Stories?
Xingmeng Zhao, Tongnian Wang, Sheri Osborn and Anthony Rios 186

Byte-ranked Curriculum Learning for BabyLM Strict-small Shared Task 2023
Justin DeBenedetto 198

McGill BabyLM Shared Task Submission: The Effects of Data Formatting and Structural Biases
Ziling Cheng, Rahul Aralikatte, Ian Porada, Cesare Spinoso-Di Piano and Jackie CK Cheung 207

Mean BERTs make erratic language teachers: the effectiveness of latent bootstrapping in low-resource settings
David Samuel 221

Not all layers are equally as important: Every Layer Counts BERT
Lucas Georges Gabriel Charpentier and David Samuel 238

WhisBERT: Multimodal Text-Audio Language Modeling on 100M Words
Lukas Wolf, Klemen Kotar, Greta Tuckute, Eghbal Hosseini, Tamar I. Regev, Ethan Gotlieb Wilcox and Alexander Scott Warstadt 253

A surprisal oracle for active curriculum language modeling
Xudong Hong, Sharid Loáiciga and Asad Sayeed 259

Mmi01 at The BabyLM Challenge: Linguistically Motivated Curriculum Learning for Pretraining in Low-Resource Settings
Maggie Mi 269

Baby Llama: knowledge distillation from an ensemble of teachers trained on a small dataset with no performance penalty
Inar Timiryasov and Jean-Loup Tastet 279

BabyLM Challenge: Curriculum learning based on sentence complexity approximating language acquisition
Miyu Oba, Akari Haga, Akiyo Fukatsu and Yohei Oseki 290

Better Together: Jointly Using Masked Latent Semantic Modeling and Masked Language Modeling for Sample Efficient Pre-training
Gábor Berend 298

Lil-Bevo: Explorations of Strategies for Training Language Models in More Humanlike Ways
Venkata S Govindarajan, Juan Diego Rodriguez, Kaj Bostrom and Kyle Mahowald 308

Towards more Human-like Language Models based on Contextualizer Pretraining Strategy
Chenghao Xiao, G Thomas Hudson and Noura Al Moubayed 317

Increasing The Performance of Cognitively Inspired Data-Efficient Language Models via Implicit Structure Building
Omar Momen, David Arps and Laura Kallmeyer 327

Pre-training LLMs using human-like development data corpus
Khushi Bhardwaj, Raj Sanjay Shah and Sashank Varma 339

On the effect of curriculum learning with developmental data for grammar acquisition
Mattia Opper, J. Morrison and N. Siddharth 346

Optimizing GPT-2 Pretraining on BabyLM Corpus with Difficulty-based Sentence Reordering
Nasim Borazjanizadeh 356