Niklas Muennighoff

GitHub: Muennighoff
X: Muennighoff
Website: muennighoff.com
Email: n.muennighoff@gmail.com

Education

2024–2028 Stanford University, PhD in Computer Science
Advisors: Yejin Choi, Andrew Ng

2018–2022 Peking University [reference], Bachelor's in Business
Valedictorian (GPA: 3.83/4.00, Rank: 1/32, Graduation Speech (CN))

Select Jobs

2021–2026 AI research at varied organizations including Cursor, Hugging Face, Meta, Nvidia, Aleph Alpha & Ai2

2018–2021 Varied internships including at Japan's Space Agency (JAXA), Ernst & Young, L.E.K. Consulting & a Ryokan

2010–2018 Self-employed voice-over artist [sample] for major film productions including Game of Thrones (HBO), Frankenweenie (Disney), Pokemon, Peter Pan (Disney) & Dracula

Select Papers

Google Scholar: https://scholar.google.com/citations?user=Me0IoRMAAAAJ&hl=en

[1] Composer 2 Technical Report
54 authors including N Muennighoff
arXiv 2026, **Most used model in Cursor with millions of daily queries**

[2] s1: Simple test-time scaling
{N Muennighoff, Z Yang, W Shi, XL Li}, H Hajishirzi, L Fei-Fei, L Zettlemoyer, P Liang, E Candès, T Hashimoto
EMNLP 2025, ICLR 2025 Reasoning Workshop **Best Paper Award**, **6.7K GitHub stars**, **arXiv 2025 (Computation and Language) Most Influential Paper #2**

[3] Molmo and PixMo: Open Weights and Open Data for State-of-the-Art Multimodal Models
M Deitke, C Clark, S Lee, R Tripathi, Y Yang, JS Park, M Salehi, N Muennighoff ...
CVPR 2025 **Best Paper Honorable Mention Award**

[4] OLMoE: Open Mixture-of-Experts Language Models
N Muennighoff, L Soldaini, D Groeneveld, K Lo, J Morrison ... H Hajishirzi
ICLR 2025 **Oral**, NeurIPS 2024 ENLSP Workshop **Oral**, **iOS App**

[5] Generative Representational Instruction Tuning
N Muennighoff, H Su, L Wang, N Yang, F Wei ... D Kiela
ICLR 2025, ICLR 2024 AGI Workshop **Best Paper Award**.

[6] Aya Model: An Instruction Finetuned Open-Access Multilingual Language Model
A Üstün ... N Muennighoff, M Fadaee, J Kreutzer, S Hooker
ACL 2024 **Best Paper Award**

[7] KTO: Model Alignment as Prospect Theoretic Optimization
K Ethayarajh, W Xu, N Muennighoff, D Jurafsky, D Kiela
ICML 2024 Spotlight, **arXiv 2024 (Machine Learning) Most Influential Paper #2**

[8] OLMo: Accelerating the Science of Language Models
D Groeneveld ... N Muennighoff ... H Hajishirzi
ACL 2024 **Theme Paper Award**, **ACL 2024 Most Influential Paper #4**

[9] Dolma: an Open Corpus of Three Trillion Tokens for Language Model Pretraining Research
L Soldaini ... N Muennighoff ... K Lo
ACL 2024 **Best Resource Paper Award**, **ACL 2024 Most Influential Paper #7**

[10] OctoPack: Instruction Tuning Code Large Language Models
N Muennighoff, Q Liu, A Zebaze, Q Zheng, B Hui ... S Longpre
ICLR 2024 Spotlight

[11] Scaling Data-Constrained Language Models
N Muennighoff, AM Rush, B Barak, TL Scao, A Piktus, ... C Raffel
NeurIPS 2023 **Outstanding Paper Runner-Up Award**, JMLR 2025

[12] StarCoder: may the source be with you!
R Li, LB Allal, Y Zi, N Muennighoff ... H Vries
TMLR 2023

[13] BLOOM: A 176B-Parameter Open-Access Multilingual Language Model
{TL Scao, A Fan ... N Muennighoff ... M Mitchell, C Raffel} ...
arXiv 2022, **arXiv 2022 (Computation and Language) Most Influential Paper #6**

[14] Crosslingual Generalization through Multitask Finetuning
N Muennighoff, T Wang, L Sutawika, A Roberts, S Biderman ... C Raffel
ACL 2023, **ACL 2023 Most Influential Paper #2**

[15] MTEB: Massive Text Embedding Benchmark
N Muennighoff, N Tazi, L Magne, N Reimers
EACL 2023, **20M installs**, **3.3K GitHub stars**

[16] Vilio: State-of-the-art Visio-Linguistic Models applied to Hateful Memes
N Muennighoff
NeurIPS 2020 Competitions, **2nd/3300+ in Meta AI Hateful Memes Competition**

Select Honors

2025 CVPR Best Paper Honorable Mention for Molmo [3]; 2 Best Papers and 5 Honorable Mentions in 13,008 submissions

2025 Knight-Hennessy Scholarship covering 3y Stanford PhD Tuition + stipend + mentorship; 84 scholars selected in 8,570 applications

2025 ICLR Reasoning Workshop Best Paper Award for s1 [2]; 1 award in 181 submissions

2024 Ai2 year-end award for my work on OLMoE [4]; 2 awards among 350 people

2024 ACL Best Theme Paper Award for OLMo [8]; 1 award in 4,407 submissions

2024 ACL Best Paper Award for Aya [6]; 5 awards in 4,407 submissions

2024 ACL Best Resource Paper Award for Dolma [9]; 3 awards in 4,407 submissions

2024 Stanford School of Engineering Fellowship covering 1y Stanford PhD Tuition + stipend

2024 ICLR AGI Workshop Best Paper Award for GRIT [5]; 1 award in 40 submissions

2023 NeurIPS Outstanding Paper Runner-Up Award for Scaling Data-Constrained Language Models [11]; 2 Outstanding and 2 Runner-Up awards in 12,343 submissions

2020 Peking University Future Leaders Full Scholarship covering tuition + stipend + mentorship; Cohort of 36 students

2020 Meta Hateful Memes Challenge 2nd/3,300+ winning 25K USD + NeurIPS presentation

Languages

Fluent: **English** (TOEFL 118/120), **Japanese** (JLPT N1), **German** (native)
Advanced: **Chinese** (100/100 Peking University Chinese Placement Test), **French** (C2)

Select Invited Talks

2026 Stanford CS525 (host: Ludwig Schmidt) guest lecture; *Scaling Data-Constrained LMs* [11]

2025 NeurIPS 2025 Workshop on Efficient Reasoning ; *Test-time scaling*

2025 UC Berkeley (host: Alexandros G. Dimakis); *s1* [2]

2025 Microsoft GenAI (host: Shuohang Wang); *s1: Simple test-time scaling* [2]

2024 Stanford NLP (host: Chris Manning); *OLMoE*, *Dolma*, *OLMo*, *Molmo* [4, 9, 8, 3]

2024 Cohere For AI; *Scaling Data-Constrained Language Models* [11]

2024 Amazon (host: Rana Ali Amjad); *GRIT* [5]

2023 Univ. of Toronto (host: Colin Raffel); *Scaling Data-Constrained LMs* [11]

2022 Princeton Univ. (hosts: Mengzhou Xia, Danqi Chen); *Scaling Data-Constrained LMs* [11]

2022 Cornell Tech (host: Alexander M. Rush); *Scaling Data-Constrained LMs* [11]

2020 NeurIPS 2020 Hateful Memes Competition; *Vilio* [16]

Select Press

2025 TWIML AI Podcast: *s1: A High-Performance Reasoning Model* [2]

2025 DeepLearning.AI: *s1: Simple test-time scaling* [2]

2025 TechCrunch: *[s1] an open rival to OpenAI's o1* [2]

2024 Economic Report of the President of the USA citing *Scaling Data-Constrained LMs* [11]

2022 Samuel Albanie: *Crosslingual generalization through multitask finetuning* [14]

2022 Peking University News: *Niklas Muennighoff: Positively Impacting Humanity*