

Maarten Bosma

LLM POST-TRAINING & EVALUATION

✉ maarten@bosma.de | 🏠 ma2rten.github.io | 📧 ma2rten | 🌐 maartenbosma

Summary

AI researcher specializing in LLM post-training and evaluation. Co-author of *Chain-of-Thought* and *Emergent Abilities*; initiator of instruction tuning with *FLAN* (ICLR oral); creator of the *MBPP* program-synthesis dataset; and core contributor to *PaLM*. Founding team at Inflection AI (*Inflection-2*; second-best LLM at release), working across pretraining data pipelines, finetuning, safety, and evaluation. At Microsoft, worked to adopt GPT-4o real-time voice in Copilot.

Experience

Microsoft AI

PRINCIPAL MEMBER OF TECHNICAL STAFF

Mountain View, CA

Mar 2024 – Oct 2024

- Worked at Microsoft and on-site at OpenAI to launch and adopt GPT-4o real-time voice in Copilot.
- Contributed evaluation planning and acceptance criteria for adoption, including latency and robustness considerations.

Inflection AI

MEMBER OF TECHNICAL STAFF — FOUNDING TEAM

Palo Alto, CA

Apr 2022 – Mar 2024

- We trained *Inflection-2* (second-best LLM at time of release) and released *Pi*.
- Created early pretraining data pipelines (ingestion, dedup/filtering, quality) and supporting infra; trained/evaluated early model families.
- Worked across post-training, safety, and evaluation (instruction tuning, evaluation approaches).

Google

RESEARCH ENGINEER (GOOGLE BRAIN)

Mountain View, CA

Aug 2016 – Apr 2022

- Initiated *FLAN* (ICLR oral); co-authored *Chain-of-Thought* and *Emergent Abilities*.
- Core contributor to *PaLM*; created *MBPP* to assess program-synthesis ability of large LMs.
- Additional work on dialog systems and on pretraining data impacts on safety/quality; earlier work applying deep learning in Google Ads.

Whisper

DATA SCIENTIST / MACHINE LEARNING ENGINEER

Venice, CA

Jan 2014 – Aug 2016

- Improved engagement by more than 10% via personalized feeds; platform served ~30M MAU.
- Built a background-image suggestion system used by over 50% of creators; implemented high-quality post identification.
- Used deep learning (LSTMs) and traditional NLP (topic modeling, POS tagging, entity extraction) with MapReduce/Hadoop/Hive over large-scale datasets.

Worknet

SOFTWARE DEVELOPER

Remote

Jul 2012 – Dec 2013

- Implemented full-text CV search and a mobile web/app while working remotely.

Hyves

SOFTWARE ENGINEER (INTERN & PART-TIME)

Amsterdam, NL

Jan 2011 – Jan 2012

- Bachelor's thesis on adult image detection (metadata ensemble + Bag-of-Visual-Words/SIFT); anti-abuse/anti-spam tooling; one publication from this work.

Selected Publications

Finetuned Language Models are Zero-Shot Learners (FLAN)

WEI, BOSMA, ZHAO, GUU, YU, LESTER, DU, DAI, LE

2022

International Conference on Learning Representations (ICLR) — Oral

Chain-of-Thought Prompting Elicits Reasoning in Large Language Models

WEI, WANG, SCHUURMANS, BOSMA, ICHTER, XIA, CHI, LE, ZHOU

2022

Advances in Neural Information Processing Systems (NeurIPS)

Emergent Abilities of Large Language Models

WEI, TAY, BOMMASANI, RAFFEL, **BOSMA**, ZHOU, METZLER, CHI, HASHIMOTO, VINYALS, LIANG, DEAN, FEDUS
Transactions on Machine Learning Research (TMLR)

2022

PaLM: Scaling Language Modeling with Pathways

CHOWDHERY, NARANG, DEVLIN, **BOSMA**, *et al.*
Journal of Machine Learning Research (JMLR), 24

2023

Program Synthesis with Large Language Models (introduces MBPP)

AUSTIN, ODENA, NYE, **BOSMA**, *et al.*
ICLR 2022 Workshop

2022

Full list: Google Scholar.

Education

MSc Studies, Artificial Intelligence (coursework)

UNIVERSITY OF AMSTERDAM
Machine Learning and Natural Language Processing.

Netherlands

2010 – 2012

BSc, Computer Science

UNIVERSITY OF AMSTERDAM
Minor in International Development Studies.

Netherlands

2007 – 2010

Skills

Post-Training	Instruction tuning, RLHF/RLAIF, safety, red teaming
Evaluation	Human-feedback data pipelines, reward modeling, evaluation approaches
Stack	Python, C++, PyTorch, JAX/TF, TensorFlow; Azure/GCP; Apache Spark, Hadoop/Hive/MapReduce, BigQuery
Languages	English (Full Professional), Dutch/German (Native/Bilingual), Thai (Conversational)