Marc van Zee

Staff Research Engineer

+45 (44) 16 46 15
marcvanzee@gmail.com
marcvanzee.github.io
marcvanzee
marcvanzee
OPZa8z4AAAAJ

Professional Summary

Staff Research Engineer with over 12 years of research and industry experience in language and image/video generation. Led the development of optimized attention mechanisms at DeepMind, reducing training and inference time by up to 40% for industry-leading models like Imagen 3 and Veo. Previously, as Tech Lead at Google Brain, scaled Flax's adoption from 100s to 10,000s users with 0 team expansion, enabling its use in flagship projects like Gemini, AlphaFold, and Vision Transformers. Collaborative leader passionate about driving innovation and fostering inclusive, high-performing teams.

Professional Experience

Jun 2022 – Present — **Google DeepMind**, Copenhagen, DK

Staff Research Engineer (Oct 2024 – Present)

Senior Research Engineer (Oct 2022 - Sep 2024)

- I played a key role in Veo, developing a text-to-video model for Google I/O 2024. My contributions included extensive code refactoring, enhancements to the input pipeline, critical bug fixes, and innovative improvements to the architecture, leading to the project's highlighted success at the event.
- As an Imagen 3 core contributor, I led the implementation of an advanced attention kernel, enabling attention at higher resolutions than previously possible. My attention module became integral to numerous DeepMind models.
- I significantly improved productivity by migrating the Imagen 3 training code to a new modular Flax codebase. It was adopted as the foundation for the Image team, resulting in all Imagen 3 models and training code utilizing it.

Jun 2016 – Sep 2022 — **Google**

Research Engineer, Brain, Amsterdam, NL (Nov 2019 - May 2022)

- As a key contributor and co-Tech Lead of Flax, I led the team to grow Flax from a niche tool with hundreds of users to the primary neural network library used across GDM by tens of thousands of users, powering high-impact projects like Gemini, Alphafold and T5X. I led weekly syncs, set team priorities, and made significant technical contributions, remaining the #3 all-time code contributor, providing leadership during rapid growth and team shrinkage.

Software Engineer, Research, Zurich, CH (Feb 2017 - Oct 2019)

- Key developer of the CFQ dataset for compositional generalization in language, co-authoring a highly cited ICLR 2020 paper (>1,000 citations) and leading a follow-up paper as first author (>100 citations).

PhD SWE Internship, Shopping, Pittsburgh, USA (Jun 2016 – Aug 2016)

- Developed a constraint satisfaction resolution algorithm for resolving product attribute conflicts in C++.

Technical Skills

Programming Languages	**Proficient:** Python, C++ **Familiar:** Java, JavaScript, C, Bash, Go
Frameworks & Libraries	**Deep Learning Frameworks:** JAX (including Flax), PyTorch, TensorFlow **Libraries & Tools:** NumPy, SciPy, Scikit-learn
Areas of Expertise	**Machine Learning:** Deep Learning, Large-scale Model Training, Model Parallelism, Distributed Computing **Natural Language Processing:** Language Analysis and Processing, Language Generation Models **Computer Vision:** Image Generation, Video Generation, Diffusion Models **Optimization Techniques:** Neural Network Optimization, Attention Mechanisms, Model Scaling
Other Skills	High-Performance Computing, Cross-Functional Team Leadership, Agile Methodologies, Open-Source Development

Education

2014 – 2017 — **Ph.D. in Computer Science**, *University of Luxembourg*, Luxembourg

- **Thesis**: "Rational Architecture: Architecture from a Recommender Perspective"
- **Achievements**:
  - Graduated with a grade of **Outstanding (5/5)**
  - Exchange program at **Stanford University**, enhancing research through international collaboration

2011 – 2013 — **M.Sc. in Artificial Intelligence**, *Utrecht University*, Utrecht, Netherlands

- **Achievements**:
  - Graduated **Summa Cum Laude** with a GPA of **9.6/10**
  - Thesis project on **UAVs** at **Linköping University**, Sweden, focusing on advanced AI applications in robotics

2005 – 2009 **B.Sc. in Industrial Design**, *Eindhoven University of Technology*, Eindhoven, Netherlands

- **Achievements**:
  - Participant in the **Honours Programme**, demonstrating academic excellence and commitment to advanced learning

Publications and Patents

Full list of publications available at: Google Scholar Profile

Patent

- **van Zee, M.**, Fruchter, S., Donahue, J., Chen, Y., Gupta, A., Chung, J. "Video Generation and Encoding Using Machine-Learning Models." Patent Application IDF-01871, 2024 (pending).

Selected Publications

1. **Baldridge, J.**, Bauer, J., Bhutani, M., Brichtova, N., Bunner, A., **van Zee, M.**, et al. (2024). "Imagen 3." *arXiv preprint* arXiv:2408.07009.
2. **van Zee, A.**, **van Zee, M.**, Søgaard, A. (2024). "Group Fairness in Multilingual Speech Recognition Models." In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics* (NAACL), pages 2213–2226.
3. **Roberts, A.**, Chung, H. W., Levskaya, A., Mishra, G., Bradbury, J., Andor, D., Narang, S., Lester, B., Gaffney, C., Mohiuddin, A., Hawthorne, C., Lewkowycz, A., Salcianu, A., **van Zee, M.**, et al. (2023). "Scaling Up Models and Data with T5X and SeqIO." *Journal of Machine Learning Research*, 24(377):1–8.
4. **Keysers, D.**, Schärli, N., Scales, N., Buisman, H., Furrer, D., Kashubin, S., Momchev, N., Sinopalnikov, D., Stafiniak, L., Tihon, T., Tsarkov, D., Wang, X., **van Zee, M.**, Bousquet, O. (2020). "Measuring Compositional Generalization: A Comprehensive Method on Realistic Data." In *Proceedings of the International Conference on Learning Representations* (ICLR).
5. **van Zee, M.**, Furrer, D., Scales, N., Schärli, N. (2020). "Compositional Generalization in Semantic Parsing: Pre-training vs. Specialized Architectures." *arXiv preprint* arXiv:2007.08970.
6. **Heek, J.**, Levskaya, A., Oliver, A., Ritter, M., Rondepierre, B., Steiner, A. and **van Zee, M.**, et al. (2020). "Flax: A Neural Network Library and Ecosystem for JAX." Github repository

Honors and Professional Activities

2024 **Imagen 3 and Veo Featured at Google I/O 2024**
Presented by DeepMind's CEO Demis Hassabis as highlights of Google I/O.

2022 **Research Tech Impact Award for Flax**
Recognized for significant contributions to the Flax library.

2021–2024 **Google Recognition Awards**
Received 8 spot bonuses and 9 peer bonuses for outstanding performance.

2014–2016 **Board Member, Benelux AI Association (BNVKI)**
Served on the board, contributing to AI community initiatives.

2013–2024 **Reviewer for Premier AI Conferences**
Reviewed over 65 papers for ICLR, ICML, AAAI, NeurIPS, IJCAI, etc.

Open Source Contributions

- **Flax.** *Tech Lead and Core Contributor; #3 contributor on GitHub.* Neural network library built on JAX. Integral to projects like Gemini, AlphaFold, and Vision Transformers.
- **JAX on the Web.** *Sole Author.* Library for converting JAX models to TFLite (mobile) and TF.js (web).
- **T5X.** *Core Contributor.* Modular framework for high-performance training and inference of sequence models.
- **Flax for HuggingFace.** *Sole Author.* Developed the first Flax models (Tweeted by Jeff Dean)
- **Additional Projects.** Numerous other projects available on GitHub.

Leadership and Mentorship

- **Mentoring Junior Engineers**
  Mentored 2 junior engineers during the Flax project, fostering skill development and professional growth.
- **Internship Supervision**
  Hosted intern Roxana Pop during COVID-19. Received exceptional feedback for mentorship.
- **Technical Interviewer at Google**
  Conducted over 70 software engineering coding interviews, contributing to team growth and talent acquisition.
- **Technical Presentations**
  Delivered over 60 presentations and deep dives at various reading groups and team meetings.
- **Coach for Google's "Tech Introduction" Project**
  Led training sessions for new joiners and organized an open forum for training other coaches globally.