Hanna Tseran, Ph.D.

MACHINE LEARNING RESEARCH SCIENTIST

Tokyo, Japan (Open to relocation | Visa sponsorship required) • hanna.tseran@gmail.com • +81-90-7736-2572
hanna-tseran.github.io • Google Scholar • linkedin.com/in/hanna-tseran • github.com/hanna-tseran

SUMMARY

Machine learning researcher focusing on **deep learning theory** with a goal to establish fundamental principles underlying modern neural networks, enabling more capable and explainable models. Currently, I am particularly fascinated by the mechanisms of **optimization** in large **transformer-based models**, and by the representations and **emergent abilities** that arise as a result of training. I have published in **NeurIPS**, **ICML**, and **TMLR**, and worked on machine learning research at RIKEN AIP, Amazon, and Microsoft Research. My long-term goal is to advance mathematically grounded understanding of deep learning systems to improve the architecture and optimization of foundation models.

SKILLS

Research Areas: Deep Learning Theory, Foundation Models, Large Language Models (LLMs), Transformers, Emergent Abilities, In-Context Learning, Optimization
Technical Skills: Python, PyTorch, TensorFlow, Parallel Computing (MPI), HPC (Slurm, UGE)

EXPERIENCE

RIKEN Center for Advanced Intelligence Project (AIP) — Tokyo, Japan
Postdoctoral Researcher — *Apr 2024 – present*
High-Dimensional Structure Theory Team | Team leader: Prof. Masaaki Imaizumi

- Lead several lines of research on **LLM behavior and capabilities**.
- Proved the existence of a transformer architecture capable of learning-to-learn optimal autoregressive process estimators, offering a theoretical explanation for **in-context learning**; preprint forthcoming.
- Theoretically established saddle-to-saddle dynamics in **transformer optimization** under relaxed architectural assumptions, offering insight into generalization in neural networks; preprint in preparation.
- Secured a JSPS Grant-in-Aid for Early-Career Scientists to develop a theoretical framework for in-context learning based on singular learning theory, aiming to further the understanding of **emergent abilities in LLMs**.
- Engage in student supervision and collaboration within the joint RIKEN–University of Tokyo lab and beyond.

The University of Tokyo — Tokyo, Japan
Project Researcher — *Nov 2023 – Mar 2024*
Matsuo Lab

- Explored theoretical directions in **LLMs**, including work on **training dynamics** now continued at RIKEN.

Max Planck Institute for Mathematics in the Sciences (MPI MiS) — Leipzig, Germany
Ph.D. Student — *Jan 2020 – Nov 2023*
Mathematical Machine Learning Team

- Proposed a **stable initialization method** for deep maxout networks, achieving over **40% accuracy improvement**; published in ICML.
- Proved that **expected functional complexity** grows polynomially with depth in maxout networks, correcting earlier exponential assumption and providing **guidance on network size scaling**; published in NeurIPS.
- Contributed to a study on loss landscapes of ReLU networks; published in TMLR.

Amazon — Berlin, Germany
Applied Scientist Intern — *Nov 2022 – Mar 2023*
Natural Language Processing Team

- Developed an **efficient memory-augmented transformer** architecture for conversational AI, enabling processing of sequences of unbounded length.

Microsoft Research   Cambridge, UK
Research Software Engineer   *Dec 2018 – Dec 2019*

TrueSkill project used by Halo, the game selling 81MM+ copies. A team of 5 people, research led by Dr. Tom Minka
- Developed components of the player skill estimation system based on Bayesian networks, **translating machine learning research to practical applications**.
- Explored system applicability to alternative game designs, obtaining quantitative results.
- Quickly ramped up in C# to enhance metrics for performance analysis and improve library design.

RIKEN Center for Advanced Intelligence Project (AIP)   Tokyo, Japan
Research Assistant   *Nov 2017 – Aug 2018*

Approximate Bayesian Inference Team. Team leader: Dr. M. Emtiyaz Khan
- Derived a novel continual learning method based on the approximate variational inference algorithm for Bayesian neural networks and wrote a paper accepted to a NeurIPS workshop.

Google   Dublin, Ireland
Site Reliability Engineering (SRE) Intern   *Jul 2017 – Oct 2017*
- Prototyped a deep learning system for identifying spam in SRE alerts based on anomaly detection techniques.

Yandex   Minsk, Belarus
Software Engineer   *Dec 2014 – Mar 2016*

Backend team of around 20 people working on the Yandex search engine written mostly in C++, the most popular search engine in Russia at the time, with 100MM+ daily queries
- Accelerated loading and reduced memory consumption of the search engine.
- Designed and implemented an approach to optimize data center balancing.
- Discovered and implemented a method to reduce search database size without information loss, saving approximately $1M in storage costs.

EDUCATION

Max Planck Institute for Mathematics in the Sciences (MPI MiS)   Leipzig, Germany
Ph.D. in Computer Science (theoretical focus, mathematical machine learning)   *Jan 2020 – Nov 2023*

Thesis: Expected Complexity and Gradients of Deep Maxout Neural Networks and Implications to Parameter Initialization. Supervisor: Prof. Guido Montúfar (Group Leader at MPI MiS and Professor at UCLA).
Degree awarded by Leipzig University. Research conducted at MPI MiS with parallel enrollment in IMPRS.

The University of Tokyo   Tokyo, Japan
Master of Information Science and Technology   *Sep 2016 – Sep 2018*

Thesis: Variational Inference for Continual Learning by using Weight-Perturbation in Adam. Supervisor: Prof. Tatsuya Harada.

Belarusian State University   Minsk, Belarus
Specialist Degree in Computer Science   *Sep 2010 – Jun 2015*

Thesis: Algorithms for recognition of circular objects and elements on them (in case of coins). Supervisor: Prof. Yuri Svirid.

GRANTS & SCHOLARSHIPS

JSPS Grant-in-Aid for Early-Career Scientists for 4,810,000¥ ($\approx$33,000$) | Principal Investigator | Topic: Theoretical Framework for In-Context Learning Development in LLMs Based on Singular Learning Theory | Awarded Apr 2025
Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) Scholarship, one of two recipients from Belarus | Full tuition and stipend | Apr 2016 – Aug 2018

SELECTED PUBLICATIONS

Karhadkar, Kedar and Murray, Michael and **Tseran, Hanna** and Montúfar, Guido. **Mildly Overparameterized ReLU Networks Have a Favorable Loss Landscape.** *Transactions of Machine Learning Research, TMLR* (2024)
Tseran, Hanna, and Montúfar, Guido. **Expected Gradients of Maxout Networks and Consequences to Parameter Initialization.** *International Conference on Machine Learning, ICML* (2023)
Tseran, Hanna, and Montúfar, Guido. **On the Expected Complexity of Maxout Networks.** *Advances in Neural Information Processing Systems, NeurIPS* (2021)