

# HANNA TSERAN

## Researcher in Deep Learning Theory



Tokyo, Japan



hanna-tseran.github.io



hanna.tseran@gmail.com



Ru1jQaAAAAAJ



hanna-tseran



hanna-tseran

## Summary

Researcher in deep learning with **diverse machine learning expertise** and **industry experience**. I am broadly interested in machine learning and particularly fascinated by **deep learning theory** for its benefits for the informed design of deep learning components. In academia, I research **implicit bias** in neural networks, analyzed the neural network **expressivity** and **initialization**, obtained results on the expected effects of the **activation function** choice, and developed algorithms for **continual** and **few-shot learning**. In the industry, I designed a **memory-augmented transformer** for conversational AI during the Amazon internship. Previously, I enhanced the engineering design and researched the transferability of a **Bayesian network** model to new game settings in Microsoft. I am curious about various network architectures and have worked with **transformers**, **Bayesian neural networks**, **VAEs**, **GANs**, and **GNNs**. I believe **combining theoretical and applied knowledge** is the best way to advance the machine learning field.

## Research Interests

mathematical machine learning

deep learning

theory

optimization

large language models (LLMs)

expressivity

expected network behavior

implicit bias

architectures: transformer, GNN

complex setups: zero-shot, continual learning

applications and theory interplay

## Programming Skills

Python

PyTorch, TensorFlow

Parallel computing, MPI

C#, C++

## Experience

### ► Postdoctoral Researcher

RIKEN Center for Advanced Intelligence Project (AIP)

Apr 2024 – present

Tokyo, Japan

High-Dimensional Causal Analysis Team. Team leader: Prof. Masaaki Imaizumi

- Deep learning theory, especially in application for Large Language Models

### ► Project Researcher

The University of Tokyo

Nov 2023 – Mar 2024

Matsuo Lab

Tokyo, Japan

- Deep learning theory for Large Language Models

### ► Ph.D. Student in Computer Science

Max Planck Institute for Mathematics in the Sciences (MPI MiS) Leipzig, Germany

Jan 2020 – Jul 2023

Mathematical Machine Learning group. Thesis: Expected Complexity and Gradients of Deep Maxout

Neural Networks and Implications to Parameter Initialization. Supervisor: Prof. Guido Montúfar

- Investigated effects of the activation function choice (maxout activation) on the expected network gradients, derived consequences to parameter initialization, expressivity, and NTK, and published an ICML paper
- Analyzed the activation function (maxout activation) influence on the expected complexity of neural networks, discovered that it does not grow exponentially with the network depth in maxout networks, and published a NeurIPS paper
- Researched independently and collaborated with other researchers in and outside the lab on the topics of algebraic geometry methods in deep learning, neural network Lipschitz constant, graph neural network NTK

### ► Applied Scientist Intern

Amazon

Nov 2022 – Mar 2023

Natural Language Processing

Berlin, Germany

- Designed an efficient transformer augmented with external memory to allow for unbounded context in conversational AI (publication in preparation)

### ► Research Software Engineer

Microsoft Research

Dec 2018 – Dec 2019

TrueSkill project for estimating player skill based on Bayesian networks; used by Halo, the game selling

Cambridge, UK

81MM+ copies. A team of 5 people, research led by Dr. Tom Minka

- Researched ways to transfer the model to a game with a different design and obtained preliminary results on the model applicability
- Enhanced model metrics to facilitate the performance analysis: adjusted existing and added new ones; improved library design, e.g., class structure
- Learned a new language (C#) in a month, interviewed engineering candidates, attended Machine Learning Summer School

### ► Research Assistant

RIKEN Center for Advanced Intelligence Project

Nov 2017 – Aug 2018

Approximate Bayesian Inference Team. Supervisor: Dr. M. Emteyaz Khan

Tokyo, Japan

- Developed a novel continual learning method based on the Bayesian deep learning algorithm and wrote a paper accepted to a NeurIPS workshop

### ► Site Reliability Engineering (SRE) Intern

Google

Jul 2017 – Oct 2017

Dublin, Ireland

- Researched and implemented the proof-of-concept spam detection system based on anomaly detection using deep learning to monitor SRE messages for a system used by Google SREs

### ► Research Student

The University of Tokyo

Apr 2016 – Sep 2016

Yasuo Kuniyoshi's Laboratory

Tokyo, Japan

- Studied emotion recognition in images and implemented a baseline solution

### ► Software Engineer

Yandex

Dec 2014 – Mar 2016

Backend team of around 20 people working on the Yandex search engine, the most popular search engine in Russia at that moment, with 100MM+ daily queries

Minsk, Belarus

- Accelerated loading and reduced memory consumption of the search
- Designed and implemented an approach to optimize data center balancing

### ► Software Engineering Intern

Yandex

Jul 2014 – Aug 2014

Moscow, Russia

- Discovered and implemented a method reducing the size of a search database without the loss of information saving around 1MM\$ in storage

# HANNA TSERAN

## Researcher in Deep Learning Theory



Tokyo, Japan



hanna-tseran.github.io



hanna.tseran@gmail.com



Ru1jQaAAAAAJ



hanna-tseran



hanna-tseran

## Reviewing Duties

ICML 2022, TMLR

## Scholarships

Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) Scholarship, one of two winners in Belarus. Apr 2016 – Aug 2018

## Language Skills

English Proficient

Japanese Upper-Intermediate, JLPT N2

## Education

- **Ph.D. in Computer Science** Jan 2020 – Nov 2023  
Max Planck Institute for Mathematics in the Sciences (MPI MiS) and International Max Planck Research School (IMPRS MiS).  
The degree was awarded by Leipzig University Leipzig, Germany  
Mathematical Machine Learning group. Thesis: Expected Complexity and Gradients of Deep Maxout Neural Networks and Implications to Parameter Initialization. Supervisor: Prof. Guido Montúfar
- **Master of Information Science and Technology** Sep 2016 – Sep 2018  
The University of Tokyo Tokyo, Japan  
Machine Intelligence lab. Thesis: Variational Inference for Continual Learning by using Weight-Perturbation in Adam. Supervisor: Prof. Tatsuya Harada
- **Specialist Degree in Computer Science** Sep 2010 – Jun 2015  
Belarusian State University Minsk, Belarus  
Thesis: Algorithms for recognition of circular objects and elements on them (in case of coins). Supervisor: Prof. Yuri Svirid

## Publications

- Tseran, Hanna. **Expected Complexity and Gradients of Deep Maxout Neural Networks and Implications to Parameter Initialization.** *Doctoral Thesis, Supervisor: Prof. Guido Montúfar* (2023)
- Tseran, Hanna, and Cheng, Wang. **Turing Machine Transformer for Unbounded Sequence Processing.** *Under review* (2023)
- Karhadkar, Kedar and Murray, Michael and Tseran, Hanna and Montúfar, Guido. **Mildly Overparameterized ReLU Networks Have a Favorable Loss Landscape.** *arXiv preprint arXiv:2305.19510* (2023)
- 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)
- Tseran, Hanna and Khan, Mohammad Emtiyaz and Harada, Tatsuya and Bui, Thang D. **Natural Variational Continual Learning.** *NeurIPS Workshop on Continual Learning* (2018)
- Tseran, Hanna. **Variational Inference for Continual Learning by using Weight-Perturbation in Adam.** *Master's Thesis, Supervisor: Prof. Tatsuya Harada* (2018)
- Tseran, Hanna, and Harada, Tatsuya. **Memory augmented neural network with Gaussian embeddings for one-shot learning.** *NeurIPS Workshop on Bayesian Deep Learning* (2017)
- Tseran, Hanna. **Algorithms for recognition of circular objects and elements on them in case of coins** (In Russian). *Specialist Thesis, Supervisor: Prof. Yuri Svirid* (2015)

## Selected Talks

- Regular talks on different topics several times per semester. *Deep Learning Theory Seminar, MPI MiS & UCLA* (2020 – present)
- Expected Gradients of Maxout Networks and Consequences to Parameter Initialization.**, Poster talk. *ICML* (2023)
- Expected Complexity and Gradients of Maxout Networks**, 15 minute talk. *Annual Meeting of the Theoretical Foundations of Deep Learning Program, Evangelische Akademie in Tutzing, Germany* (2022)
- Expected Complexity and Gradients of Maxout Networks**, 1 hour invited talk. *Nonlinear Algebra Seminar, MPI MiS, Germany* (2022)
- Expected Complexity of Maxout Networks**, 1 hour invited talk. *AI + Math Colloquia at INS, Shanghai Jiao Tong University (SJTU)* (2022)
- On the Expected Complexity of Maxout Networks**, Poster talk. <https://slideslive.com/38967795>. *NeurIPS* (2021)