# Róbert Csordás

e-mail: rcsordas@openai.com website: https://robertcsordas.github.io/

## EDUCATION

**USI/IDSIA** http://idsia.ch 2018 April - 2023 September
PhD Lugano, Switzerland
Supervised by Prof. Jürgen Schmidhuber.
Thesis: **Systematic Generalization in Connectionist Models**.
https://sonar.ch/documents/326205/files/2023INF013.pdf

**Budapest University of Technology and Economics** Budapest, Hungary
Electrical Engineering. MSc (grad. 2015) and BSc (grad. 2012). Grade: excellent.

## WORK EXPERIENCE

**OpenAI** https://openai.com/ 2025 July - present
Member of Technical Staff (Research Scientist) London, United Kingdom
Working on architectures for future large language models.

**Stanford** https://stanford.edu/ 2024 February - 2025 July
Postdoctoral Researcher Stanford, California, USA
I am working on systematic generalization and improving language models. Supervised by Prof. Christopher Manning and Prof. Christopher Potts.

**IDSIA** http://idsia.ch 2023 October - 2024 January
Postdoctoral Researcher Lugano, Switzerland
I worked on systematic generalization.

**DeepMind** - https://www.deepmind.com 2022 June - 2022 October
Research Scientist Intern London, United Kingdom
I worked on graph neural networks, improving generalization on algorithmic problems and external memory for Transformers.

**AImotive (formerly AdasWorks)** - https://aimotive.com 2015-2018
AI Research Scientist Budapest, Hungary
I worked on deep neural networks for self driving cars.

- Monocular depth prediction using neural networks.
- Neural stereo matching - predicting robust depth map with neural network.
- Recurrent network research - Convolutional LSTMs for stabilizing detections, free space detection, etc.
- Object detection, semantic segmentation

**Hungarian Academy of Sciences - Institute for Computer Science and Control** - https://www.sztaki.hu 2015
Software Engineer Budapest, Hungary
I worked on classical computer vision projects. For example:

- Detecting objects thrown over the fence; detecting human leaving a car.
- Autonomous forklift control system.

**Innomed Medical Inc.** - http://innomed.hu 2007 - 2015
Embedded Software/Hardware Engineer Budapest, Hungary

- Designed the software architecture of Linux based patient monitor (C++, QT).

- Maintained the software of the InnoCare-S patient monitor (C++).
- Wrote low level hardware drivers for InnoCare-T12.

**PUBLICATIONS**

Houjun Liu, Shikhar Murty, Christopher D. Manning, Róbert Csordás: **Thought-bubbles: an Unsupervised Method for Parallel Thinking in Latent Space** - We propose an unsupervised method for parallel thinking in latent space, by learning to fork residual streams in the Transformer.
*arXiv preprint* `https://arxiv.org/abs/2510.00219`

Anand Gopalakrishnan, Róbert Csordás, Jürgen Schmidhuber, Michael C. Mozer: **Decoupling the" What" and" Where" With Polar Coordinate Positional Embeddings** - We propose a new, complex positional encoding that outperforms RoPE and generalizes better to longer sequences.
*arXiv preprint* `https://arxiv.org/abs/2509.10534`

Róbert Csordás, Christopher D. Manning, Christopher Potts: **Do Language Models Use Their Depth Efficiently?** - We analyze the residual stream of popular LLMs and show that they do not use their depth efficiently and they are not composiontal.
*NeurIPS 2024* `https://arxiv.org/abs/2505.13898`

Aryaman Arora, Neil Rathi, Nikil Roashan Selvam, Róbert Csordás, Dan Jurafsky, Christopher Potts: **Mechanistic evaluation of Transformers and state space models** - We analyze Transformers and SSMs on associative recall tasks and show that they are using fundamentally different mechanisms.
*arXiv preprint* `https://arxiv.org/abs/2505.15105`

Joakim Edin, Róbert Csordás, Tuukka Ruotsalo, Zhengxuan Wu, Maria Maistro, Jing Huang, Lars Maaløe: **GIM: Improved Interpretability for Large Language Models** - We propose a better attribution method for the attention layers.
*arXiv preprint* `https://arxiv.org/abs/2505.17630`

Piotr Piękos, Róbert Csordás, Jürgen Schmidhuber: **Mixture of Sparse Attention: Content-Based Learnable Sparse Attention via Expert-Choice Routing** - We propose a novel sparse attention mechanism with expert-choice routing.
*arXiv preprint* `https://arxiv.org/abs/2505.00315`

Vincent Herrmann, Róbert Csordás, Jürgen Schmidhuber: **Measuring In-Context Computation Complexity via Hidden State Prediction** - We propose a novel method to quantify the "interestingness" of in-context computation in neural networks.
*ICML 2025* `https://arxiv.org/abs/2503.13431`

Julie Kallini, Shikhar Murty, Christopher D. Manning, Christopher Potts, Róbert Csordás: **MrT5: Dynamic Token Merging for Efficient Byte-level Language Models** - We propose a dynamic token deletion mechanism that forces merging information in the encoder of ByT5, speeding it up significantly.
*ICLR 2025* `https://arxiv.org/abs/2410.20771`

Róbert Csordás, Christopher Potts, Christopher D. Manning, Atticus Geiger: **Recurrent Neural Networks Learn to Store and Generate Sequences using Non-Linear Representations** - We found evidence of RNNs storing sequences in a nonlinear way.

*BlackboxNLP 2024* https://arxiv.org/abs/2408.10920

Róbert Csordás, Kazuki Irie, Jürgen Schmidhuber, Christopher Potts, Christopher D. Manning: **MoEUT: Mixture-of-Experts Universal Transformers** - We propose an MoE Universal Transformer that works well on large-scale language modeling tasks for the first time.
*NeurIPS 2024* https://arxiv.org/abs/2405.16039

Róbert Csordás, Piotr Piękos, Kazuki Irie, Jürgen Schmidhuber: **SwitchHead: Accelerating Transformers with Mixture-of-Experts Attention** - We present an MoE attention that can match the performance of parameter-matched dense models.
*NeurIPS 2024* https://arxiv.org/abs/2312.07987

Kazuki Irie, Róbert Csordás, Jürgen Schmidhuber: **Metalearning Continual Learning Algorithms** - We propose a self-referential neural network to meta-learn its own in-context continual learning algorithms.
*TMLR 2025* https://arxiv.org/abs/2312.00276

Róbert Csordás, Kazuki Irie, Jürgen Schmidhuber: **Approximating Two-Layer Feedforward Networks for Efficient Transformers** - We present an improved MoE that can match the performance of parameter-matched dense models.
*EMNLP Findings 2023* https://arxiv.org/abs/2310.10837

Mingchen Zhuge, Haozhe Liu, Francesco Faccio, Dylan R. Ashley, Róbert Csordás, Anand Gopalakrishnan, Abdullah Hamdi, Hasan Abed Al Kader Hammoud, Vincent Herrmann, Kazuki Irie, Louis Kirsch, Bing Li, Guohao Li, Shuming Liu, Jinjie Mai, Piotr Piękos, Aditya Ramesh, Imanol Schlag, Weimin Shi, Aleksandar Stanić, Wenyi Wang, Yuhui Wang, Mengmeng Xu, Deng-Ping Fan, Bernard Ghanem, Jürgen Schmidhuber **Mindstorms in Natural Language-Based Societies of Mind** - We discuss ideas on multi-agent LLM systems.
*arXiv preprint* https://arxiv.org/abs/2305.17066

Kazuki Irie, Róbert Csordás, Jürgen Schmidhuber: **Practical Computational Power of Linear Transformers and Their Recurrent and Self-Referential Extensions** - We show that linear transformers inherit several capabilities from standard Transformers, and self-referential extensions successfully overcome some of them.
*EMNLP 2023* https://aclanthology.org/2023.emnlp-main.588/

Kazuki Irie*, Róbert Csordás*, Jürgen Schmidhuber: **Topological Neural Discrete Representation Learning à la Kohonen** - We show that VQ used in VQ-VAEs is a special case of SOMs, which are more robust and converge faster.
*ICANN 2024* https://arxiv.org/abs/2302.07950

Anian Ruoss, Grégoire Delétang, Tim Genewein, Jordi Grau-Moya, Róbert Csordás, Mehdi Bennani, Shane Legg, Joel Veness: **Randomized Positional Encodings Boost Length Generalization of Transformers** - We propose randomized, ordered positional encodings to improve length generalization on algorithmic tasks.
*ACL 2023* https://arxiv.org/abs/2305.16843

Róbert Csordás, Kazuki Irie, Jürgen Schmidhuber: **CTL++: Evaluating Generalization on Never-Seen Compositional Patterns of Known Functions, and Compatibility of Neural Representations** - We extend the CTL dataset to test systematicity and show how NNs develop incompatible representations and fail to generalize.

*EMNLP 2022* https://arxiv.org/abs/2210.06350

Borja Ibarz, Vitaly Kurin, George Papamakarios, Kyriacos Nikiforou, Mehdi Bennani, Róbert Csordás, Andrew Dudzik, Matko Bošnjak, Alex Vitvitskyi, Yulia Rubanova, Andreea Deac, Beatrice Bevilacqua, Yaroslav Ganin, Charles Blundell, Petar Veličković: **A Generalist Neural Algorithmic Learner** - We show that graph neural networks can learn many algorithms together and generalize to larger problem instances.
*LoG 2022* https://arxiv.org/abs/2209.11142

Kazuki Irie*, Róbert Csordás*, Jürgen Schmidhuber: **The Dual Form of Neural Networks Revisited: Connecting Test Time Predictions to Training Patterns via Spotlights of Attention** - We investigate dual-form representations of NNs to understand how their behavior depends on the training samples.
*ICML 2022* https://arxiv.org/abs/2202.05798

Kazuki Irie, Imanol Schlag, Róbert Csordás, Jürgen Schmidhuber: **A Modern Self-Referential Weight Matrix That Learns to Modify Itself** - We propose a scalable self-referential layer that uses self-generated training patterns, outer products, and the delta update rule to modify itself.
*ICML 2022* https://arxiv.org/abs/2202.05780

Róbert Csordás, Kazuki Irie, Jürgen Schmidhuber: **The Neural Data Router: Adaptive Control Flow in Transformers Improves Systematic Generalization** - We propose to improve data routing in Transformers by gating and geometric attention, achieving systematic generalization on algorithmic tasks.
*ICLR 2022* https://arxiv.org/abs/2110.07732

Róbert Csordás, Kazuki Irie, Jürgen Schmidhuber: **The Devil is in the Detail: Simple Tricks Improve Systematic Generalization of Transformers** - We significantly improve the systematic generalization of Transformers on various systematic generalization datasets using simple tricks.
*EMNLP 2021* https://arxiv.org/abs/2108.12284

Kazuki Irie, Imanol Schlag, Róbert Csordás, Jürgen Schmidhuber: **Going Beyond Linear Transformers with Recurrent Fast Weight Programmers** - We explore the recurrent Fast Weight Programmers (FWPs), which exhibit advantageous properties of both Transformers and RNNs.
*NeurIPS 2021* https://arxiv.org/abs/2106.06295

Kazuki Irie, Imanol Schlag, Róbert Csordás, Jürgen Schmidhuber: **Improving Baselines in the Wild**
*NeurIPS 2021 DistShift* https://openreview.net/forum?id=9vxOrkNTs1x

Róbert Csordás, Sjoerd van Steenkiste, Jürgen Schmidhuber: **Are Neural Nets Modular? Inspecting Functional Modularity Through Differentiable Weight Masks** - We develop a method for analyzing emerging functional modularity in neural networks based on differentiable weight masks and use it to point out important issues in current-day neural networks.
*ICLR 2021* https://openreview.net/forum?id=7uVcpu-gMD

Róbert Csordás, Jürgen Schmidhuber: **Improving Differentiable Neural Computers Through Memory Masking, De-allocation, and Link Distribution Sharpness Control** - Addresses 3 different issues with the original DNC architec-

ture. Also proposes a new, better content-based lookup mechanism.
*ICLR 2019* `https://openreview.net/forum?id=HyGEM3C9KQ`

Róbert Csordás, László Havasi, and Tamás Szirányi: **Detecting objects thrown over fence in outdoor scenes** - A new technique for detecting objects thrown over a critical area of interest in a video sequence made by a monocular camera.
*VISAPP 2015* `http://goo.gl/ZDkk4g`

**GRANTS**

**CSCS** - My proposal, "Improving Systematic Generalization of Neural Networks", won 250000 GPU-hours on the Piz Daint supercomputer. 2022

**PATENTS**

Róbert Csordás, Ágnes Kis-Benedek, Balázs Szalkai: **Method and Apparatus for Generating a Displacement Map of an Input Dataset Pair** - A neural network based method for fast and robust stereo matching for depth map generation.
*US10380753* `https://patents.google.com/patent/US10380753B1`

**ORGANIZING WORKSHOPS**

**System 2 Reasoning At Scale** at NeurIPS 2024
`https://s2r-at-scale-workshop.github.io/`

**TEACHING EXPERIENCE**

**Research Projects** - Working with PhD students and supervising undergrad projects.
Stanford 2024
**Semester Project Cosupervision** - Analyzing emergence
ETH Zürich 2023
**Teaching Assistant** - Deep Learning Lab
Università della Svizzera italiana 2021, 2022
**Teaching Assistant** - Machine Learning
Università della Svizzera italiana 2018, 2019, 2020

**HIGH SCHOOL PUBLICATIONS**

**CallTheTux** - CallTheTux, a universal GSM stack for Linux.
*Petnica Papers, 2007* `https://goo.gl/QTCy5U`
**RealVM** - A new type of virtual machine that would allow parallel execution and fast switching between different operating systems.
*Petnica Papers, 2006* `https://goo.gl/8TNHf5`
**PrologAPI** - Enabling the usage of Prolog constructs from C++.
*Petnica Papers, 2005* `https://goo.gl/KpV3sF`

**TECHNICAL STRENGTHS**

Python, PyTorch, TensorFlow, C, C++, CUDA, OpenCV, Algorithms, Linux, JavaScript, Bash, Matlab, Assembly

**OTHER SKILLS**

**Machine learning frameworks:** PyTorch, JAX, TensorFlow, Torch
**Parallel programming:** CUDA, Triton, Numba
**Electronics:** KiCAD, Eagle, PIC, PIC32, AVR, AVR32, ARM, XMOS, Xilinx
**Databases:** MySQL, MongoDB, Sphinx search
**JavaScript technologies:** NodeJS, jQuery
**Mobile development:** Android, iOS (Swift)
**Operating systems:** Linux, OS X, Windows
**Markup languages:** LaTeX, XML, Markdown

**HOBBY PROJECTS**

**MobilECG II** - `https://github.com/robertcsordas/MobilECG-II` 2014 - 2016
Open source Holter ECG. Designed the schematic diagram and the firmware.

**engineerjs.com** - `http://engineerjs.com` 2013 - 2015
Extendable online computing environment for engineers, with physical quantity, complex numbers and linear algebra support.

**LANGUAGES**

Hungarian (native); English, Serbian (fluent); Italian (intermediate); German (beginner)