

Mariya Toneva

+49 (151) 21159581
mtoneva@mpi-sws.org
<http://mtoneva.com>

Positions Held

Max Planck Institute for Software Systems

Tenure-track Faculty (W2)

since 2022

Visiting Researcher

2021–2022

Max Planck School of Cognition

Adjunct Faculty

since 2023

Princeton University

C.V. Starr Fellow

2021–2022

Mentors: Ken Norman, Uri Hasson

Education

Carnegie Mellon University

Ph.D. in Machine Learning and Neural Computation

2014–2021

Thesis title: Bridging Language in Machines with Language in the Brain

Advisors: Tom Mitchell, Leila Wehbe

Carnegie Mellon University

Masters of Science in Machine Learning

2018

Yale University

Bachelor of Science in Computer Science, Cognitive Science

2014

Publications in Journals and Conference Proceedings

1. Improving semantic understanding in speech language models via brain-tuning
O. Moussa, D. Klakow, and **M. Toneva**
(ICLR 2025) *International Conference on Learning Representations* [pdf] 2025
2. Large language models can segment narrative events similarly to humans
S. Michelmann, M. Kumar, K.A. Norman, and **M. Toneva**
Behavioral Research Methods [pdf] 2025
3. Hints help finding and fixing bugs differently in python and text-based
program representations
R. Rawal, V. Padurean, S. Apel, A. Singla, and **M. Toneva**
(ICSE 2025) *International Conference on Software Engineering* [pdf] 2025
4. Language models and brains align due to more than next-word prediction and
word-level information
G. Merlin and **M. Toneva**
(EMNLP 2024) *Empirical Methods in Natural Language Processing* [pdf] 2024

5. Speech language models lack important brain-relevant semantics

S.R. Oota, E. Çelik, F. Deniz, and **M. Toneva**

(ACL 2024) *Annual Meeting of the Association for Computational Linguistics* <http://aclanthology.org/2024>

6. Perturbed examples reveal invariances shared by language models

R. Rawal and **M. Toneva**

(ACL Findings 2024) *Annual Meeting of the Association for Computational Linguistics* [\[pdf\]](#)

7. Joint processing of linguistic properties in brains and language models

S.R. Oota, M. Gupta, and **M. Toneva**

(NeurIPS 2023) *Neural Information Processing Systems* [\[pdf\]](#)

8. What happens during finetuning of vision Transformers: an invariance based investigation

G. Merlin, V. Nanda, R. Rawal, and **M. Toneva**

(CoLLAs 2023) *Conference on Lifelong Learning Agents* [\[pdf\]](#)

9. Training language models for deeper understanding improves brain alignment

K.L. Aw and **M. Toneva**

(ICLR 2023) *International Conference on Learning Representations* [\[pdf\]](#)

[\[top 25% notable paper \(Spotlight\)\]](#)

10. A Roadmap to reverse engineering real-world generalization by combining naturalistic

paradigms, deep sampling, and predictive computational models

P. Herholz, E. Fortier, **M. Toneva**, N. Farrugia, L. Wehbe, V. Borghesani

Neurons, Behavior, Data Science, and Theory [\[pdf\]](#)

11. Combining computational controls with natural text reveals aspects of

meaning composition

M. Toneva, T. Mitchell, and L. Wehbe

Nature Computational Science [\[pdf\]](#)

12. Same cause; different effects in the brain

M. Toneva*, and J. Williams*, A. Bollu, C. Dann, and L. Wehbe

(CLEaR 2022) *Causal Learning and Reasoning* [\[pdf\]](#)

13. Single-trial MEG data can be denoised through cross-subject predictive modeling

S. Ravishankar, **M. Toneva**, and L. Wehbe

Frontiers in Computational Neuroscience 2021 [\[pdf\]](#)

14. Modeling task effects on meaning representation in the brain

via Zero-Shot MEG Prediction

M. Toneva*, O. Stretcu*, B. Poczoz, L. Wehbe, and T. Mitchell

(NeurIPS 2020) *Neural Information Processing Systems* [\[pdf\]](#)

15. Interpreting and improving natural-language processing (in machines) with

natural language-processing (in the brain)

M. Toneva and L. Wehbe

(NeurIPS 2019) *Neural Information Processing Systems* [\[pdf\]](#)

17. An empirical study of example forgetting during deep neural network learning 2019
M. Toneva*, A. Sordoni*, R. Tachet des Combes*, A. Trischler, Y. Bengio, and G. Gordon
 (ICLR 2019) *International Conference on Learning Representations* [\[pdf\]](#)
18. Applying artificial vision models to human scene understanding 2015
 E. M. Aminoff, **M. Toneva**, A. Shrivastava, X. Chen, I. Misra, A. Gupta, and M. J. Tarr
Frontiers in Computational Neuroscience 2015 [\[pdf\]](#)
19. Exploration of social grouping: effects of behavioral mimicry, appearance, and eye gaze 2014
 A. Nawroj, **M. Toneva**, H. Admoni, B. Scassellati
 (CogSci 2014) *Conference of the Cognitive Science Society* [\[with Oral presentation\]](#) [\[pdf\]](#)
20. The physical presence of a robot tutor increases cognitive learning gains 2012
 D. Leyzberg, S. Spaulding, **M. Toneva**, and B. Scassellati
 (CogSci 2012) *Conference of the Cognitive Science Society* [\[pdf\]](#)
21. Robot gaze does not reflexively cue human attention 2011
 H. Admoni, C. Bank, J. Tan, **M. Toneva**, and B. Scassellati
 (CogSci 2011) *Conference of the Cognitive Science Society* [\[pdf\]](#)

Preprints and Non-Proceeding Publications

- Position: episodic memory is the missing piece for long-term LLM agents 2025
 M. Pink, Q. Wu, V. Vo, J. Turek, J. Mu, A. Huth, and **M. Toneva**
 (arXiv 2025) [\[pdf\]](#)
- Assessing episodic memory in LLMs with sequence order recall tasks 2024
 M. Pink, V. Vo, ... and **M. Toneva**
 (arXiv 2024) [\[pdf\]](#)
- Vision-language integration in multimodal video transformers (partially) aligns with the brain 2023
 D. Dong and **M. Toneva**
 (arXiv 2023) [\[pdf\]](#)
- Getting aligned on representational alignment 2023
 I. Sucholutsky, L. Muttenthaler, ..., **M. Toneva**, T. Griffiths
 (arXiv 2023) [\[pdf\]](#)
- Pointwise representational similarity 2023
 C. Kolling, T. Speicher, V. Nanda, **M. Toneva**, and K.P. Gummadi
 (arXiv 2023) [\[pdf\]](#)
- Interpreting multimodal video Transformers using brain recordings 2023
 D. Dong and **M. Toneva**
 (ICLR 2023 Workshop on Multimodal Representation Learning: Perks and Pitfalls) [\[pdf\]](#)

- Memory for long narratives

M. Toneva, V. Vo, J. Turek, S. Jain, S. Michelmann, M. Capotă, A. Huth, U. Hasson, and K. Norman

(CEMS 2022) *Context and Episodic Memory Symposium*

2022
- The Courtois Neuromod project: a deep, multi-domain fMRI dataset to build individual brain models

J. Boyle*, B. Pinsard*, ... **M. Toneva**, ... and P. Bellec

(HBM 2022) *Human Brain Mapping* [with Oral presentation]

2022
- Does injecting linguistic structure into language models lead to better alignment with brain recordings?

M. Abdou, A.V. González, **M. Toneva**, D. Hershcovich, and A. Søgaard

(arXiv 2021) [\[pdf\]](#)

2021
- Investigating different alignment methods between natural and artificial neural networks for language processing

A. Bollu, **M. Toneva**, and L. Wehbe

(SNL 2020) *Society for the Neurobiology of Language*

2020
- Investigating task effects on brain activity during stimulus presentation in MEG

M. Toneva*, O. Stretcu*, B. Poczos, and T. Mitchell

(HBM 2019) *Human Brain Mapping*

2019
- Word length processing in left lateraloccipital through region-to-region connectivity: an MEG Study

M. Toneva, and T. Mitchell

(HBM 2018) *Human Brain Mapping*

2018
- MEG representational similarity analysis implicates hierarchical integration in sentence processing

N. Rafidi*, D. Schwartz*, **M. Toneva***, S. Jat, and T. Mitchell

(HBM 2018) *Human Brain Mapping*

2018
- Scene-space encoding within the functional scene-selective network

E. M. Aminoff, **M. Toneva**, A. Gupta, and M. J. Tarr

(VSS 2015) *Vision Sciences Society*

2015
- Towards a model for mid-level feature representation of scenes

M. Toneva, E. M. Aminoff, A. Gupta, and M. Tarr

(VSS 2014) *Vision Sciences Society*

2014

Grants and Fellowships

DFG Graduate School RTG 2853

Neuroexplicit Models of Language, Vision & Action

2023-2028

DFG Research Unit 5368 KI-FOR

Abstract Representations in Neural Architectures

Project: Bridging Levels of Abstraction in Brains and Natural Language Processing Machines

2023-2027

National Institutes of Health T32 Training Grant	
Funded one year of postdoctoral research in Quantitative and Computational Neuroscience	2021–2022
C.V. Starr Fellowship	
Funded one year of postdoctoral research in computational neuroscience at Princeton University	2021–2022
National Science Foundation Graduate Research Fellowship	
Funded three years of interdisciplinary graduate research in machine learning and neuroscience	2016–2019
Grace Hopper Celebration Scholarship	
Funded attendance at the 2014 Grace Hopper Celebration of Women in Computing	2014
Mellon Forum Undergraduate Research Grant	
Funded submission and attendance at the 2014 Vision Sciences Society conference	2014
Robin Berlin Fellowship	
Funded neural modeling research at Laboratory of Computational Neuroscience, EPFL	2013

Awards and Distinctions

Japanese-American-German Frontiers of Science (JAGFOS) Symposium Invitee	
Alexander von Humboldt Foundation	2023
Selected to represent Germany as one of 24 researchers from all disciplines of science and engineering	
Appointed as Member	
ELLIS (European Laboratory for Learning and Intelligent Systems)	2023
Appointed as Adjunct Fellow	
Max Planck School of Cognition	2023
Ph.D. Dissertation Award, Honorable Mention	
Society for the Neurobiology of Language	2021
Machine Learning Student Leadership Award	
Awarded for exemplary efforts and their significant impact on life in the Machine Learning Department	2020
Top Reviewer	
NeurIPS 2022, ICLR 2022, ICML 2022, NeurIPS 2018	2018–2022
Citadel Datathon Runner-up	
Analyzed a genomics dataset to predict age-related differences in disease-related gene expression	2017
Machine Learning Teaching Assistant Award	
Awarded for outstanding performance as a TA in 10-725 Convex Optimization	2017
BrainHub Neurohackathon Winner	
Reduced need for human supervision by classifying diffusion MRI tracks into anatomical bundles	2016

Invited and Contributed Talks

Conferences and Workshops

▪ Keynote Speaker , CALCULUS Symposium, KU Leuven	2024
▪ Invited Speaker, Brain Prize Webinar “LLMs and Human Language Processing”	2024
▪ Keynote Speaker , AAAI Workshop “AI for Brain Encoding and Decoding”	2024
▪ Invited Speaker, ICLR Workshop “Representational Alignment”	2024
▪ Invited Speaker, CogSci Symposium “Is DL the Answer for Understanding Cog. Dynamics”	2024
▪ Invited Speaker, German Academy of Science Symposium “Brain Science and LLMs”	2024
▪ Invited Speaker, NeuroAI Symposium	2024
▪ Keynote Speaker , Neuro-AI Talks (NEAT)	2023
▪ Invited Speaker, Workshop on Philosophy of Science Meets ML, University of Tübingen	2023
▪ Invited Symposium Speaker, European Society for Philosophy and Psychology Conference	2023
▪ Invited Speaker, Neuro-AI Educational Session, Human Brain Mapping Conference	2023
▪ Invited Speaker, Workshop on Code, Brains, and LLMs, Saarland University	2023
▪ Invited Speaker, Workshop on ML, Abstract Thought, and the Expanding Reach of AI	2022
▪ Contributed Talk, Neuromatch Conference	2020
▪ Invited Speaker, CCN Workshop on Nonlinear Models for Scientific Discovery	2020
▪ Contributed Talk, NeurIPS Workshop Women in Machine Learning	2014
▪ Contributed Talk, CogSci Oral Presentation for Accepted Paper	2014

University Seminars

▪ Institute Colloquium, Gatsby Computational Neuroscience Unit	2025
▪ Institute Colloquium, Ernst Strüngmann Institut	2024
▪ Math Machine Learning Seminar, Max Planck Institute for Mathematics and UCLA	2024
▪ Institute Colloquium, MPI for Human Cognitive and Brain Sciences	2024
▪ Campus Lecture, Saarbrücken Informatics Campus	2023
▪ Cog. Comp. Neuro. Colloquium, MPI for Human Cognitive and Brain Sciences	2023
▪ Psychology Department Colloquium, University of Saarland	2023
▪ Institute Colloquium, Institute for Basic Science, South Korea	2023
▪ Institute Colloquium, CIMEC, University of Trento	2022
▪ Distinguished Speakers in Language Science Colloquium, Saarland University	2022
▪ Department of Statistics and Data Science, Yale University	2021
▪ Faculty of Computing and Data Science, Boston University	2021
▪ Department of Computer Science, University of Southern California	2021
▪ Department of Computer Science, University of Utah	2021
▪ Institute Colloquium, Max Planck Institute for Software Systems	2021
▪ Department of Computer Science, Aarhus University	2021
▪ Institute Colloquium, Toyota Technological Institute at Chicago	2021
▪ Institute Colloquium, IST Austria	2021
▪ Department of Statistics, University of Chicago	2021
▪ Department of Computer Science, University of Notre Dame	2021
▪ Department of Computer Science, University of Liverpool	2020
▪ SFB-TRR 161 Lecture Series, U. of Stuttgart, U. of Konstanz, Ulm University, and LMU	2020

Summer Schools and Group Meetings

▪ Heinrich Heine University Düsseldorf, Host: Milica Gasic	2024
▪ Goethe University Frankfurt, Host: Christian Fiebach	2024
▪ Bernstein Center for Computational Neuroscience Retreat	2024
▪ Max Planck School of Cognition Academy	2023
▪ Summer School in Philosophy and Computer Science, University of Bayreuth	2023
▪ Department of Computer Vision and Machine Learning, MPI for Informatics	2023
▪ TALEP group, Aix-Marseille University	2022
▪ CMMRS Summer School, MPI for Software Systems	2022
▪ IMPRS NeuroCom Summer School at the MPI for Human Cognitive and Brain Sciences	2022
▪ MIT, Host: Evelina Fedorenko	2022
▪ MIT, Host: Roger Levy	2021
▪ Computational Neuroscience Symposium, CMU	2021
▪ Courtois NeuroMod Group, University of Montreal	2021
▪ Princeton Neuroscience Institute, Hosts: Ken Norman and Uri Hasson	2021
▪ brAln seminar, CMU	2020
▪ UT Austin, Host: Alexander Huth	2020

Industry Internships

Microsoft Research, Montreal

Research Intern 2018
Investigated the learning dynamics of neural networks as they train on single classification tasks, finding that certain examples are forgotten with high frequency, and some not at all, and that, based on these forgetting dynamics, a significant fraction of examples can be omitted from the training data set while still maintaining state-of-the-art generalization performance

Cognitive Computing Center, Thomson Reuters

Research Intern 2017
Investigated the use of a recurrent neural network encoder for unsupervised word-order sensitive hashing as a step towards improving ranking results

Research Visits

Carnegie Mellon University

Research Assistant; Advisor: Michael Tarr 2013–2014
Investigated mid-level scene representation in humans using computer vision techniques

École Polytechnique Fédérale de Lausanne (EPFL)

Summer Intern; Advisor: Wulfram Gerstner 2013
Worked towards improving the state-of-the-art calcium-based model of spike-timing dependent plasticity

Massachusetts Institute of Technology

Technical Trainee; Advisor: John Gabrieli 2012
Examined links between working memory capacity and various brain metrics through the analysis of resting state functional connectivity fMRI data

Mentorship and Supervision

Omer Moussa

PhD at University of Saarland

2024-

Camila Kolling

PhD at University of Saarland
co-advised by Krishna Gummadi

2024-

Mathis Pink

PhD at University of Saarland
co-advised by Isabel Valera

2024-

Emin Çelik

Postdoctoral Researcher

2023-

Blerta Veseli

PhD at University of Saarland
co-advised by Alexander Koller

2023-

Cameron Braunstein

PhD at University of Saarland
co-advised by Eddy Ilg

2023-

Gabriele Merlin

PhD at CS@MaxPlanck Graduate Program

2022-

Alan Sun

Research Intern; now Masters student at CMU

2024

Khai Loong Aw; now PhD student at Stanford University

Research Intern

2022

Ruchit Rawal

Research Intern; now PhD student at University of Maryland

2022-2024

Subba Reddy Oota

Research Intern; now postdoc at TU Berlin

2022

Tianai (Dota) Dong

Masters Student; now PhD student at the MPI for Psycholinguistics

2021-2022

Teaching

Bridging Language in Machines and Language in the Brain, University of Saarland

Instructor

2023

Seminar course

3370 Mathematical Neuroscience, University of Pittsburgh

Teaching Assistant

2018

10-725 Convex Optimization, CMU

Teaching Assistant

2016

Awarded Machine Learning TA award

Machine Learning for Neuroscience, Multimodal Neuroimaging Training Program

Instructor

2016

Created curriculum and instructed 4-week course; video recordings can be found on personal webpage

Service

Organizer

Deep Learning for Brain Encoding and Decoding Tutorial, IJCAI	2023
Memory in Artificial and Real Intelligence Workshop, NeurIPS	2022
Deep Learning for Brain Encoding and Decoding Tutorial, Cognitive Sciences Society	2022
What can NLP systems teach us about language in the brain? Symposium, Society for the Neurobiology of Language	2021
How can findings about the brain improve AI systems? Workshop, ICLR	2021

Program Committee

Program Chair: *CogSci* 2024

Senior Program Committee Member: *ACL Rolling Review* (2023–present), *NeurIPS* (2024–present), *ICML* (2025–present), *CCN* (2025–present), *CCN* Technical Program Committee (2022–2024)

Program Committee Member: ML: *NeurIPS* 2016–2023 (Top 30% Reviewer in 2018); *ICML* 2019–2023 (Top 10% Reviewer in 2022); *AAAI* 2020–2021, *CoLLAs* 2022, *ICLR* 2022–2024 (Highlighted Reviewer in 2023); **NLP:** *ACL* 2019–2021; *NAACL* 2019–2021; *EMNLP* 2020–2021; *CoNLL* 2020–2021; *AACL-IJCNLP* 2020; *EACL* 2021

Journal Reviewer: *TMLR*, *Nature Human Behavior*, *Nature Communications*; *Communications Biology*; *TICS*, *Communications of the ACM*, *Frontiers in Computational Neuroscience*

Thesis Committee Member

Viktor Kewenig (UCL, 2025); Till Speicher (MPI Software Systems, 2025); Carina Kauf (MIT, 2024); RJ Antonello (UT Austin, 2024); Julien Dirani (NYU, 2024); Bernhard Schäfl (Johannes Kepler University, 2024); Damián Pascual (ETH Zurich, 2022);

ML@CMU Blog

Chief Editor and Co-founder 2018–2020
Oversaw more than 30 research posts featuring recent ML research across 6 departments in the School of Computer Science as well as other CMU schools and departments, and more than 10 educational posts

University Leadership Student Advisory Council

Member 2015–2017
Advising senior leadership at Carnegie Mellon University on the strategic priorities of the university

Graduate Student Assembly

Representative for the Program of Neural Computation 2015–2018
Advocating for the needs of graduate students

Yale Review of Undergraduate Research in Psychology

Chief Editor 2013–2014
Reviewed 50 submissions from 31 universities, and edited 9 submissions for publication

Personal

Languages	Bulgarian (Native), English (Fluent), German (Intermediate)
Citizenship	United States, Bulgaria
Github profile	http://github.com/mtoneva
Google Scholar profile	https://scholar.google.com/citations?user=a61sk-4AAAAJ