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
Postdoctoral Researcher 2021–2022
Advisors: 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. Large language models can segment narrative events similarly to humans 2025
   S. Michelmann, M. Kumar, K.A. Norman, and **M. Toneva**
   (Accepted, Behavioral Research Methods 2025) [pdf]

2. Language models and brains align due to more than next-word prediction and word-level information 2024
   G. Merlin and **M. Toneva**
   (EMNLP 2024) [pdf]

3. Speech language models lack important brain-relevant semantics 2024
   S.R. Oota, E. Çelik, F. Deniz, and **M. Toneva**
   (ACL 2024) [pdf]

4. Perturbed examples reveal invariances shared by language models 2024
   R. Rawal and **M. Toneva**
   (ACL Findings 2024) [pdf]

5. Joint processing of linguistic properties in brains and language models 2023
S.R. Oota, M. Gupta, and **M. Toneva**
(NeurIPS 2023) *Neural Information Processing Systems* [pdf]

6. What happens during finetuning of vision Transformers: an invariance based investigation 2023
G. Merlin, V. Nanda, R. Rawal, and **M. Toneva**
(CoLLAs 2023) *Conference on Lifelong Learning Agents* [pdf]

7. Training language models for deeper understanding improves brain alignment 2023
K.L. Aw and **M. Toneva**
(ICLR 2023) *International Conference on Learning Representations* [pdf]
[top 25% notable paper (Spotlight)]

8. A Roadmap to reverse engineering real-world generalization by combining naturalistic 2023
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]

9. Combining computational controls with natural text reveals aspects of 2022
Meaning Composition
M. Toneva, T. Mitchell, and L. Wehbe
Nature Computational Science [pdf]

10. Same Cause; Different Effects in the Brain 2022
M. Toneva*, and J. Williams*, A. Bollu, C. Dann, and L. Wehbe
(CLeaR 2022) *Causal Learning and Reasoning* [pdf]

11. Single-trial MEG Data Can Be Denoised Through Cross-Subject Predictive Modeling 2021
S. Ravishankar, **M. Toneva**, and L. Wehbe
Frontiers in Computational Neuroscience 2021 [pdf]

12. Modeling Task Effects on Meaning Representation in the Brain 2020
via Zero-Shot MEG Prediction
M. Toneva*, O. Stretcu*, B. Poczos, L. Wehbe, and T. Mitchell
(NeurIPS 2020) *Neural Information Processing Systems* [pdf]

13. Interpreting and Improving Natural-Language Processing (in Machines) with 2019
Natural Language-Processing (in the Brain)
M. Toneva and L. Wehbe
(NeurIPS 2019) *Neural Information Processing Systems* [pdf]

14. Inducing Brain-relevant Bias in Natural Language Processing Models 2019
D. Schwartz, **M. Toneva**, and L. Wehbe
(NeurIPS 2019) *Neural Information Processing Systems* [pdf]

15. 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]

16. 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]

17. 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]

18. 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]

19. 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

- Improving semantic understanding in speech language models via brain-tuning 2024
O. Moussa, D. Klakow, and **M. Toneva**
(arXiv 2024) [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 2022
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*

- The Courtois Neuromod project: a deep, multi-domain fMRI dataset to build individual brain models 2022
J. Boyle*, B. Pinsard*, ... **M. Toneva**, ... and P. Bellec
(HBM 2022) *Human Brain Mapping* [with Oral presentation]

- Does Injecting Linguistic Structure into Language Models Lead to Better Alignment with Brain Recordings? 2021
  M. Abdou, A.V. González, **M. Toneva**, D. Hershcovich, and A. Søgaard
  (arXiv 2021) [pdf]

- Investigating Different Alignment Methods Between Natural and Artificial Neural Networks for Language Processing 2020
  A. Bollu, **M. Toneva**, and L. Wehbe
  (SNL 2020) *Society for the Neurobiology of Language*

- Investigating Task Effects on Brain Activity During Stimulus Presentation in MEG 2019
  M. Toneva*, O.Stretcu*, B. Poczos, and T. Mitchell
  (HBM 2019) *Human Brain Mapping*

- Word Length Processing in Left Lateraloccipital through Region-to-Region Connectivity: an MEG Study 2018
  M. Toneva, and T. Mitchell
  (HBM 2018) *Human Brain Mapping*

- MEG Representational Similarity Analysis Implicates Hierarchical Integration in Sentence Processing 2018
  N. Rafidi*, D. Schwartz*, **M. Toneva***, S. Jat, and T. Mitchell
  (HBM 2018) *Human Brain Mapping*

- Scene-Space Encoding within the Functional Scene-Selective Network 2015
  E. M. Aminoff, **M. Toneva**, A. Gupta, and M. J. Tarr
  (VSS 2015) *Vision Sciences Society*

- Towards a Model for Mid-level Feature Representation of Scenes 2014
  M. Toneva, E. M. Aminoff, A. Gupta, and M. Tarr
  (VSS 2014) *Vision Sciences Society*

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 2023-2027
Project: Bridging Levels of Abstraction in Brains and Natural Language Processing Machines

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

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

- **Language Modeling beyond Language Modeling** 2024
  Keynote Speaker, CALCULUS Symposium, KU Leuven
- **Tracking Information Processing in the Human Brain** 2024
  Keynote Speaker, Workshop on Artificial Intelligence for Brain Encoding and Decoding, AAAI
- **Language Modeling beyond Language Modeling** 2023
  Keynote Speaker, Neuro-AI Talks (NEAT)
- **Why Do Large Language Models Align with Human Brains?** 2023
  Workshop on Philosophy of Science Meets Machine Learning
- **Language Modeling beyond Language Modeling** 2023
  Invited Symposium Speaker, European Society for Philosophy and Psychology (ESPP)
- **Language Modeling beyond Language Modeling** 2023
  Neuro-AI Educational Session, Organization for Human Brain Mapping (OHBM)
- **Tracking Information Processing in the Human Brain** 2023
  Workshop on Code, Brains, and LLMs

- **Deep Neural Networks as Model Organisms for Human Language Comprehension** 2022
  ML, Abstract Thought, and the Expanding Reach of AI: Ethical and Conceptual Frontiers
- **Modeling Task Effects on Meaning Representation in the Brain** 2020
  Traditional Talk, *Neuromatch Conference*
- **Nonlinear Models for Scientific Discovery about Language in the Brain** 2020
  Cognitive Computational Neuroscience (CCN) workshop
  Is it that simple? The use of linear models in cognitive neuroscience
- **Towards a Model for Mid-level Feature Representation of Scenes** 2014
  Oral presentation, *Women in Machine Learning (WiML) workshop at NeurIPS*
- **Exploration of Social Grouping: Effects of Behavioral Mimicry, Appearance, and Eye Gaze** 2014
  Oral presentation, *Conference of the Cognitive Science Society (CogSci)*

University Seminars

- **Convergence and Divergence between Language Models and Human Brains** 2024
  Ernst Strüngmann Institut (ESI)
- **Convergence and Divergence between Language Models and Human Brains** 2023
  Cog. Comp. Neuro. Colloquium, MPI for Human Cognitive and Brain Sciences
- **Why Do Large Language Models Align with Human Brains?** 2023
  Campus Lecture, Saarbrücken Informatics Campus
- **NLP systems as model organisms for language processing in the human brain** 2023
  Psychology Department Colloquium, University of Saarland
- **Why Do Large Language Models Align with Human Brains?** 2023
  Institute for Basic Science, South Korea
- **Why Do Large Language Models Align with Human Brains?** 2022
  CIMeC, University of Trento; Host: Raffaella Bernardi
- **Bridging Language in Machines with Language in the Brain** 2022
  Distinguished Speakers in Language Science Colloquium, Saarland University
- **Data-Driven Direct Transfer of Insight between Brains and AI Systems** 2021
  Department of Statistics and Data Science, Yale University
- **Data-Driven Transfer of Insight between Brains and AI Systems** 2021
  Faculty of Computing and Data Science, Boston University
- **Data-Driven Transfer of Insight between Brains and AI Systems** 2021
  Department of Computer Science, University of Southern California
- **Data-Driven Transfer of Insight between Brains and AI Systems** 2021
  Department of Computer Science, University of Utah
- **Data-Driven Transfer of Insight between Brains and AI Systems** 2021
  Max Planck Institute for Software Systems

- **Data-Driven Transfer of Insight between Brains and AI Systems** 2021
  Department of Computer Science, Aarhus University
- **Data-Driven Transfer of Insight between Brains and AI Systems** 2021
  Toyota Technological Institute at Chicago
- **Data-Driven Transfer of Insight between Brains and AI Systems** 2021
  IST Austria
- **Data-Driven Transfer of Insight between Brains and AI Systems** 2021
  Department of Statistics, University of Chicago
- **Data-Driven Transfer of Insight between Brains and AI Systems** 2021
  Department of Computer Science, University of Notre Dame
- **Data-Driven Transfer of Insight between Brains and AI Systems** 2020
  Department of Computer Science, University of Liverpool
- **Data-Driven Direct Transfer of Insight between Brains and AI Systems** 2020
  SFB-TRR 161 Lecture Series (University of Stuttgart, University of Konstanz, Ulm University, and the LMU Munich), Host: Lewis Chuang

Summer Schools and Group Meetings

- **Why Do Large Language Models Align with Human Brains?** 2023
  Max Planck School of Cognition Academy
- **Bridging Language in Machines with Language in the Brain** 2023
  Summer School in Philosophy and Computer Science, University of Bayreuth
- **Tracking Information Processing in the Human Brain** 2023
  Department of Computer Vision and Machine Learning, MPI Informatics
- **Why Do Large Language Models Align with Human Brains?** 2022
  TALEP group, Aix-Marseille University; Hosts: Abdellah Fourtassi, Carlos Ramisch
- **Bridging Language in Machines with Language in the Brain** 2022
  CMMRS Summer School, MPI for Software Systems
- **NLP Systems as Model Organisms for Human Language Comprehension** 2022
  IMPRS NeuroCom Summer School at the MPI for Human Cognitive and Brain Sciences
- **Same Cause; Different Effects in the Brain** 2022
  MIT, Host: Evelina Fedorenko
- **Data-Driven Transfer of Insight between Brains and AI Systems** 2021
  MIT, Host: Roger Levy
- **NLP Systems as Model Organisms for Human Language Comprehension** 2021
  Computational Neuroscience Symposium, CMU
- **NLP Systems as Model Organisms for Human Language Comprehension** 2021
  Courtois NeuroMod Group, Host: Pierre Bellec

- **Modeling Context-Dependent Meaning Composition During Language Comprehension** 2021
  Princeton Neuroscience Institute, Hosts: Ken Norman and Uri Hasson
- **Modeling Task Effects on Meaning Representation in the Brain** 2020
  Carnegie Mellon University, brAIn seminar
- **Composition of Context- and Task-dependent Meaning** 2020
  UT Austin, Host: Alexander Huth

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

Emin Çelik
Postdoctoral Researcher 2023-

Blerta Veseli
PhD at University of Saarland 2023-
co-advised

Cameron Braunstein
PhD at University of Saarland 2023-
co-advised

Gabriele Merlin
PhD at CS@MaxPlanck Graduate Program 2022-

Ruchit Rawal
Research Intern 2022-

Subba Reddy Oota
Research Intern 2022

Khai Loong Aw
Research Intern 2022

Tianai (Dota) Dong
Masters in Language Science and Technology at University of Saarland 2021-2022

Anand Bollu
Masters at Department of Computer Science, CMU 2019-2021

Sydney Zheng
Undergraduate at Department of Computer Science, CMU 2019

Aditri Bhagirath
Undergraduate at Department of Computer Science, CMU 2019

Tara Pirnia
MD/PhD candidate, CMU and University of Pittsburgh 2015

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 2023
Tutorial, IJCAI

Memory in Artificial and Real Intelligence 2022
Workshop, NeurIPS

Deep Learning for Brain Encoding and Decoding 2022
Tutorial, Cognitive Sciences Society

What can NLP systems teach us about language in the brain? 2021
Symposium, Society for the Neurobiology of Language

How can findings about the brain improve AI systems? 2021
Workshop, ICLR

Program Committee

Program Co-Chair: *CogSci* 2024
Area Chair: *EMNLP* 2023
Reviewer: 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), TMLR
NLP: *ACL* 2019-2021; *NAACL* 2019-2021; *EMNLP* 2020-2021; *CoNLL* 2020-2021; *AACL-IJCNLP* 2020; *EACL* 2021 **Neuro:** *Nature Human Behavior*, *Nature Communications*; *Communications Biology*; *Frontiers in Computational Neuroscience*; *Society for the Neurobiology of Language* 2022; *OHBM* 2018
Other venues: TICS, *CogSci* 2021, *Communications of the ACM*

Thesis Committee Member

RJ Antonello (UT Austin, 2024); Carina Kauf (MIT, 2024); Julien Dirani (NYU, 2024); Till Speicher (MPI Software Systems, 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