Matthew L. **Leavitt, PhD**

+1 925 575 7628 | matthew@mleavitt.net | mleavitt.net

Research Interests

My primary research interest is furthering the science of deep learning—developing and conducting rigorous, empirical experiments to better understand deep learning systems—and using the resulting insights to improve the performance, efficiency, and human-interpretability of these systems.

Education

McGill University *Montréal, Québec, Canada*

PhD in Physiology *2011 - 2017*

- **Area of specialization:** Computational & systems cognitive neuroscience
- **Advisor:** Prof. Julio C. Martinez-Trujillo, MD, PhD
- **Dissertation Title:** *Network properties underlying working memory in primate prefrontal cortex*

McGill University *Montréal, Québec, Canada*

BSc Major in Neuroscience, Minor in Musical Science and Technology *2006 - 2010*

- **Focus:** Cognitive and behavioral neuroscience (Major); Digital Signal Processing, synthesis, and sampling (Minor)

Professional Experience

AI Resident *Menlo Park, California, USA*

Facebook AI Research (FAIR) *August 2019 - February 2021*

- The FAIR AI residency brings individuals with non-traditional backgrounds into AI research by pairing them with FAIR researchers to identify and solve novel AI problems.
- Led a series of projects under the supervision of **Ari Morcos** examining how human-intuitive representations in deep networks affect generalization and robustness, and the state of interpretability research, leading to three first-author manuscripts.
- Initiated and continue to collaborate on multiple projects attempting to improve transformer models for vision tasks, which have thus far generated one publication.
- Implemented vision transformers and a number of other features in **vissl**, FAIR's open-source toolbox for self-supervised learning.

Postdoctoral Research Fellow *London, Ontario, & Montréal, Québec, Canada*

University of Western Ontario - Martinez-Trujillo Cognitive Neurophsyiology Lab *July 2017 - April 2019*

- Led a research project examining the neuronal dynamics of rule learning in prefrontal cortex in macaque monkeys.

Member, Board of Advisors *Montréal, Québec, Canada*

Canadian University Software Engineering Conference (CUSEC) *2015-present*

- CUSEC (www.cusec.net) is an annual, three-day, student-run software engineering and computer science conference.
- Provide guidance and mentorship to the Conference Chairs and staff.
- Collaborate to shape the priorities of CUSEC: ensuring the longevity and integrity of the organization and its mission.

Conference Co-Chair *Montréal, Québec, Canada*

CUSEC *2014*

- Recruited and enabled a staff of 23 across 7 teams (logistics, sponsorship, speakers, design, promotions, events, and A/V).
- Coordinated hospitality for and managed relationships with ~600 attendees (500 students, 18 speakers, dozens of sponsor representatives, and staff).

Director of Speakers *Montréal, Québec, Canada*

CUSEC *2012 - 2013*

- I directed the content, and recruited and hosted the ~18 speakers for two consecutive conferences.
- Notable speakers included Alexis Ohanian, co-founder of Reddit; Bret Victor, interface designer, whose invited talk, *Inventing On Principle*, has 800k+ views on Vimeo and YouTube; and Benjamin Black, who co-conceptualized Amazon Web Services.

Skills

Quantitative & Technical

- **Research techniques for deep neural networks**
  - *Noteworthy example:* Devised a novel regularizer to examine the causal role of easily interpretable neurons in deep neural networks (Leavitt and Morcos (2021) *ICLR*).
- **Expertise in scientific and numerical programming languages including Python and MATLAB**
  - *Noteworthy example (PyTorch):* I contributed a number of features to **vissl**, FAIR's open-source toolbox for self-supervised learning, including vision transformers, new data augmentations, and gradient clipping.
- **Machine learning & statistical techniques for high-dimensional neural data**
  - *Noteworthy example:* Decoding the contents of working memory from large networks of simultaneously-recorded neurons in prefrontal cortex in macaque monkeys (Leavitt et al. (2017) *PNAS* & Leavitt et al. (2017) *Cerebral Cortex*).
- **Modeling of neural, biological, and behavioral systems**
- **Experimental design for neuroscientific, psychological, and psychophysical research**
- **Neurosurgical techniques for research in non-human primates**
- **Digital signal processing for audio and neurophysiological applications**

Communication & Organization

- **Oral, written, and graphic communication, with a particular passion for scientific and technical content**
  - *Noteworthy examples (oral):* I have received numerous awards for oral presentations given to both general audiences and other neuroscientists (see my comprehensive CV at **mleavitt.net**), and delivered conference presentations and course lectures on behalf of my supervisor.
  - *Noteworthy example (written):* Starting in 2014, I was uniquely tasked with editing every manuscript and grant proposal produced by my doctoral lab.
- **Leadership, vision, management, logistics, content production, and vendor relations**
  - *Noteworthy example:* See my **Professional Experience** with CUSEC.
- **Scientific & technical literature review**
  - *Noteworthy example:* Leavitt et al. (2017, *Trends in Neurosciences*) reviews the evidence for working memory-related neural activity across different brain regions.
- **Teaching across diverse fields (neuroscience, music production, mathematics) and pupils (elementary through university-age students, macaque monkeys)**

Creative

- **Writing and editing for comedy on stage and in print**
- **17 years experience in music production, engineering, and composition**
- **20 years experience playing drums in a variety of genres and contexts**
- **Digital, 35mm, and medium format photography**
- **Carpentry and woodworking**

Publications, Conference Presentations, and Peer-review

I published three first-author and one contributing-author deep learning research manuscripts after only ∼1.5 years in the field, and six peer-reviewed neuroscience articles (including one review article) as part of my doctoral work. I have also given dozens of oral presentations at a range of scientific conferences and seminars. See **mleavitt.net** for lay publication summaries and links to my github and Google Scholar profile. The remainder of my CV details my peer-review experience, publications, and presentations.

Scientific peer-review (ad hoc)

International Conference on Machine Learning, Nature Neuroscience, Neuron, Proceedings of the National Academy of Sciences, Nature Communications, Cerebral Cortex, The Journal of Neuroscience, Experimental Brain Research, Frontiers in Neural Circuits

Publications and Preprints

ConViT: Improving Vision Transformers with Soft Convolutional Inductive Biases
arXiv (2021)
Under review at an archival venue
S d'Ascoli, H Touvron, ML Leavitt, AS Morcos, G Biroli, L Sagun

Selectivity considered harmful: evaluating the causal impact of class selectivity in DNNs
International Conference on Learning Representations (ICLR) (2021 - Accepted)
Also appears in *ICML Workshop on Human Interpretability in Machine Learning* (2020)
ML Leavitt, AS Morcos

Towards falsifiable interpretability research
NeurIPS Workshop: ML Retrospectives, Surveys & Meta-Analyses (2020)
ML Leavitt, AS Morcos

Linking average-and worst-case perturbation robustness via class selectivity and dimensionality
arXiv (2020)
Under review at an archival venue
Also appears in *ICML workshop on Uncertainty and Robustness in Deep Learning* (2020)
ML Leavitt, AS Morcos

A normalization circuit underlying coding of spatial attention in primate lateral prefrontal cortex
eNeuro (2019)
LR Duong, ML Leavitt, F Pieper, A Sachs, JC Martinez-Trujillo

Sustained activity encoding working memories: not fully distributed
Trends in Neurosciences (2017)
ML Leavitt, D Mendoza-Halliday, JC Martinez-Trujillo

Correlated variability modifies working memory fidelity in primate prefrontal neuronal ensembles
Proceedings of the National Academy of Sciences (2017)
ML Leavitt, F Pieper, AJ Sachs, JC Martinez-Trujillo

A quadrantic bias in prefrontal representation of visual-mnemonic space
Cerebral Cortex (2017)
ML Leavitt, F Pieper, AJ Sachs, JC Martinez-Trujillo

Single-trial decoding of intended eye movement goals from lateral prefrontal cortex neural ensembles
Journal of Neurophysiology (2015)
CB Boulay, F Pieper, M Leavitt, J Martinez-Trujillo, AJ Sachs

Structure of spike count correlations reveals functional interactions between neurons in dorsolateral prefrontal cortex area 8a of behaving primates
PLoS ONE (2013)
ML Leavitt, F Pieper F, A Sachs, R Joober, JC Martinez-Trujillo

Press

Easy-to-interpret neurons may hinder learning in deep neural networks
Facebook AI Research Blog (2020)

US travel ban blocking students from presenting their research
Ars Technica (2018)

This neuroscientist's poster showed how US travel bans stifle groundbreaking research
Quartz (2018)

Conference Presentations

Talks/Lectures

2020	**Selectivity considered harmful: evaluating the causal impact of class selectivity in DNNs**, ML Leavitt, AS Morcos NeuroMatch	*Neuromatch.io*
2019	**Network properties underlying cognition in macaque LPFC**, ML Leavitt Invited talk, Blake Richards Lab	*University of Toronto, Toronto, ON*
2018	**Ensemble mechanisms of rule learning in primate prefrontal cortex**, ML Leavitt, C Boulay, LR Duong, RA Gulli, AJ Sachs, JC Martinez-Trujillo. Nanosymposium on Decision Making: Circuits and Computations	*Society for Neuroscience, San Diego, CA*
2018	**Ensemble mechanisms of rule learning in primate prefrontal cortex**, ML Leavitt Invited talk, Xiao-Jing Wang Lab	*New York University, New York, NY*
2018	**Network properties underlying cognition in LPFC**, ML Leavitt Invited talk, Roozbeh Kiani Lab	*New York University, New York, NY*
2017	**Correlated variability modifies working memory fidelity in primate prefrontal neuronal ensembles**, ML Leavitt Neuroscience and Biology Group	*MILA, Montréal, QC*
2017	**Correlated variability modifies working memory fidelity in primate prefrontal neuronal ensembles**, ML Leavitt, AJ Sachs, JC Martinez-Trujillo. Workshop on Error-based learning in short-term and episodic memory	*Computational and Systems Neuroscience (COSYNE), Snowbird, UT*
2016	**Heterogeneous effects of neuronal ensemble size, tuning, and correlation structure on the decoding of spatial working memory in dorsolateral prefrontal cortex**, ML Leavitt, AJ Sachs, JC Martinez-Trujillo. Session on Visual Memory: Neural Mechanisms	*Vision Sciences Society, St. Pete's Beach, FL*
2016	**Noise correlation structure shapes ensemble coding of working memory in prefrontal cortex**, ML Leavitt, F Pieper, AJ Sachs, JC Martinez-Trujillo. Nanosymposium on Spatial Attention and Working Memory	*Society for Neuroscience, San Diego, CA*
2015	**Correlated variability and the fidelity of prefrontal working memory representations**, ML Leavitt, AJ Sachs, JC Martinez-Trujillo. Nanosymposium on Learning and Memory	*Society for Neuroscience, Chicago, IL*
2014	**Noise correlations and coding during spatial working memory**, ML Leavitt, JC Martinez-Trujillo. Satellite Symposium on Primate Brain Circuits and Behavior	*Canadian Association for Neuroscience, Montreal, QC*
2012	**The relation between local field potentials and single units across a microelectrode array implanted in macaque dorsolateral prefrontal cortex**, AJ Sachs, KJ Miller, F Pieper, ML Leavitt, JC Martinez-Trujillo. Nanosymposium on Signal Propagation	*Society for Neuroscience, New Orleans, LA*

Posters

2020	**Selectivity considered harmful: evaluating the causal impact of class selectivity in DNNs**, ML Leavitt, AS Morcos	*BayLearn, San Francisco, CA*
2018	**Learning-related modulation of rule representation in primate prefrontal cortex ensembles**, Leavitt ML, Boulay C, Gulli RA, Duong LR, Sachs A, Martinez-Trujillo JC	*Neuroinformatics, Montréal, QC*
2017	**Prefrontal cortex ensemble activity during during associative visuomotor rule learning in primates**, Leavitt ML, Boulay C, Gulli RA, Duong LR, Sachs A, Martinez-Trujillo JC	*Society for Neuroscience, Washington, DC*
2017	**Lateral prefrontal cortex single neuron and ensemble activity during associative learning in virtually navigating monkeys**, Duong LR, Gulli RA, Corrigan BW, Leavitt ML, Doucet G, Martinez-Trujillo JC	*Society for Neuroscience, Washington, DC*
2017	**Hippocampal single neuron and ensemble activity during associative learning in virtually navigating primates**, Gulli RA, Duong LR, Corrigan BW, Doucet G, Leavitt ML, Williams S, Martinez-Trujillo JC	*Society for Neuroscience, Washington, DC*
2017	**Correlated variability modifies working memory fidelity in primate prefrontal neuronal ensembles**, Leavitt ML, Sachs AJ, Martinez-Trujillo JC	*Canadian Association for Neuroscience, Montréal, QC*
2016	**Beyond the single neuron: Ensemble coding of working memory in primate prefrontal cortex**, Leavitt ML, Sachs AJ, Martinez-Trujillo JC	*The Future of Visual Attention, Rochester, NY*
2016	**Non-selective neurons contribute information to neuronal ensembles by modifying noise correlation structure**, Leavitt ML, Sachs AJ, Martinez-Trujillo JC	*Canadian Association for Neuroscience, Toronto, ON*
2015	**Predicting decision outcomes from single realizations of lateral prefrontal cortex neuronal activity**, Boulay C, Leavitt ML, Pieper F, Martinez-Trujillo JC, Sachs A	*Society for Neuroscience, Chicago, IL*
2015	**Neural representation of spatial working memory is divided into quadrants in primate prefrontal cortex**, Leavitt ML, Sachs AJ, Martinez-Trujillo JC	*Canadian Association for Neuroscience, Vancouver, BC*
2014	**Neural tuning affects spike-rate correlations during a spatial working memory task**, Leavitt ML, Pieper F, Sachs AJ, Martinez-Trujillo JC	*Society for Neuroscience, Washington, DC*
2014	**Single-trial dorsolateral prefrontal cortex neural trajectories predict intended saccade direction**, Boulay C, Pieper F, Leavitt ML, Martinez-Trujillo JC, Sachs AJ	*Society for Neuroscience, Washington, DC*
2014	**Correlated spiking during during spatial working memory in macaque prefrontal area 8a**, Leavitt ML, Pieper F, Sachs AJ, Martinez-Trujillo JC	*Canadian Association for Neuroscience, Montreal, QC*
2013	**Anti-correlated spike rates associated with working memory activity in macaque dorsolateral prefrontal cortex**, Leavitt ML, Pieper F, Sachs AJ, Martinez-Trujillo JC	*Society for Neuroscience, San Diego, CA*
2013	**Spike count correlation variability in visual, presaccadic, and visuopresaccadic neurons of macaque dorsolateral prefrontal cortex during a working memory task**, Leavitt ML, Pieper F, Sachs AJ, Martinez-Trujillo JC	*Canadian Association for Neuroscience, Toronto, ON*
2012	**Spike rate correlations in visual, presaccadic, and visuopresaccadic neurons in area 8a of macaque prefrontal cortex during a spatial working memory task**, Leavitt ML, Pieper F, Sachs AJ, Martinez-Trujillo JC	*Society for Neuroscience, New Orleans, LA*
2012	**Spike rate correlations vary by neuron response type during working memory in macaque prefrontal area 8A**, Leavitt ML, Pieper F, Sachs AJ, Martinez-Trujillo JC	*Canadian Association for Neuroscience, Vancouver, BC*
2012	**Spike count correlations in visual, visuomotor, and motor neurons of macaque prefrontal area 8A during working memory maintenance**, Leavitt ML, Pieper F, Sachs AJ, Martinez-Trujillo JC	*Vision Science Society, Naples, FL*

2011	**Correlated activity of dorsolateral prefrontal cortex neurons during spatial working memory maintenance**, Leavitt ML, Pieper F, Sachs AJ, Martinez-Trujillo JC	*Society for Neuroscience, Washington, DC*
2011	**Spike rate correlations between primate dorsolateral prefrontal cortex neurons during a spatial working memory task**, Leavitt ML, Schneiderman M, Martinez-Trujillo JC	*Canadian Association for Neuroscience, Quebec, QC*
2011	**Spike count correlations between primate dorsolateral prefrontal cortex neurons during a spatial working memory task**, Martinez-Trujillo JC, Leavitt ML, Schneiderman M	*Vision Sciences Society, Naples, FL*