Wilka Torrico De Carvalho, *Aspiring Cognitive Scientist and Machine Learner*

| CONTACT INFORMATION | *Website:* cogscikid.com
E-mail: wcarvalho@g.harvard.edu | *Github:* github.com/wcarvalho
Google Scholar |

RESEARCH INTERESTS

My long-term goal is to develop theories for human learning and generalization. In particular, I'm interested in how we discover and exploit richly structured representations to enable sophisticated behavior and rapid adaptation. Currently, I'm interested in leveraging successor features to provide a unified framework for how brains learn, combine and transfer predictive knowledge and behavior.

EMPLOYMENT

Harvard, Boston, MA — *Sept 2023 – Present*
Research Fellow

DeepMind, London, UK — *Sept 2022 – May 2023*
Research Scientist Intern

DeepMind, London, UK — *Mar 2021 – July 2021*
Research Scientist Intern

Microsoft Research, Redmond, WA — *June 2018 – Aug 2018*
Research Scientist Intern

IBM Research, San Jose, CA — *Sept 2017 – Dec 2017*
Research Scientist Intern

EDUCATION

University of Michigan–Ann Arbor, Ann Arbor, MI — Sep 2018 - July 2023
School of Engineering, Ph.D. in Computer Science
Advisors: Satinder Singh, Honglak Lee, Richard Lewis

University of Southern California, Los Angeles, CA — Aug 2015 - May 2017
Viterbi School of Engineering, M.S. in Computer Science
Advisor: Yan Liu

Stony Brook University, Stony Brook, NY — Aug 2011 - May 2015
College of Arts and Sciences, B.S. in Physics
Advisor: Axel Drees

Brooklyn Technical High School, Brooklyn, NY — May 2007 - May 2011
Diploma in Applied Physics

HONORS & AWARDS

Kempner Institute Neuro-AI Fellowship	2023
University of Michigan "The Glass is Half-Full Award"	2019
1/200 chosen internationally for Heidelberg Laureate Forum	2018
GEM National Fellowship sponsored by IBM, Adobe	2017, 2018
University of Michigan Rackham Merit Fellowship	2017
ICLR, Neurips BAI Travel Award	2017, 2019
NSF Graduate Research Fellowship (Neuroscience)	2015
Provost Award for Academic Excellence (20/5000 graduates chosen)	2015
Researcher of the Month (1 monthly at Stony Brook University)	2014
HHMI Minority Undergraduate Research Fellowship	2014
$\Sigma\Pi\Sigma$ Physics Honor Society (sponsored by Alfred Goldhaber)	2013

PREPRINTS   **Wilka Carvalho**, Andre Saraiva, Angelos Filos, Loic Matthew, Richard L. Lewis, Honglak Lee, Satinder Singh, Daniel Zoran, Danilo Rezende. *"Deep Option Keyboard: Combining Skills in Deep Reinforcement Learning."* 2023

Wilka Carvalho, Andrew Lampinen, Kyriacos Nikiforou, Felix Hill, Murray Shanahan. *"Feature-Attending Recurrent Modules Facilitate Generalization Across Object-Centric Tasks."* 2022

CONFERENCE
PUBLICATIONS   **Wilka Carvalho**, Angelos Filos, Richard L. Lewis, Honglak Lee, Satinder Singh. *"Composing Task Knowledge with Modular Successor Feature Approximators."* In ICLR, 2023

Lajanugen Logeswaran, **Wilka Carvalho**, Honglak Lee. *"Learning Compositional Tasks from Language Instructions."* In AAAI, 2023

Chris Hoang, Sungryull Sohn, Jongwook Choi, **Wilka Carvalho**, Honglak Lee. *"Successor Feature Landmarks for Long-Horizon Goal-Conditioned Reinforcement Learning."* In NeurIPS, 2021

Wilka Carvalho, Anthony Liang, Kimin Lee, Sungryull Sohn, Honglak Lee, Richard L. Lewis, Satinder Singh. *"Reinforcement Learning for Sparse-Reward Object-Interaction Tasks in First-person Simulated 3D Environments."* In IJCAI, 2021

Wilka Carvalho*, Sanjay Purushotham*, Tanachat Nilanon, Yan Liu. *"Variational Recurrent Adversarial Domain Adaptation."* In ICLR, 2017

WORKSHOP
PUBLICATIONS   Lajanugen Logeswaran, **Wilka Carvalho**, Honglak Lee. *"Learning Compositional Tasks from Language Instructions."* In NeurIPS Deep RL Workshop, 2021

Wilka Carvalho, Murray Shanahan. *"Learning to Represent State with Perceptual Schemata."* In ICML {Unsupervised RL, Real Life RL} Workshops, 2021

Wilka Carvalho, Anthony Liang, Kimin Lee, Sungryull Sohn, Honglak Lee, Richard L. Lewis, Satinder Singh. *"Reinforcement Learning for Sparse-Reward Object-Interaction Tasks in First-person Simulated 3D Environments."* In {ICML Object-Oriented Learning, NeurIPS DeepRL} Workshops, 2020

Chris Hoang, Sungryull Sohn, Jongwook Choi, **Wilka Carvalho**, Honglak Lee. *"Successor Landmarks for Efficient Exploration and Long-Horizon Navigation."* In NeurIPS DeepRL Workshop, 2020

Wilka Carvalho, Kimin Lee, Anthony Liang, Ryan Krueger, Richard L. Lewis, Satinder Singh, Honglak Lee. *"Efficiently Learning to Perform Household Task with Object-oriented Exploration."* In NeurIPS BAI, 2019 **(Oral)**

Bryant Chen*, **Wilka Carvalho***, Benjamin Edwards, Taesung Lee, Ian Molloy, Heiko Ludwig. *"Detecting Backdoor Attacks on Deep Neural Networks by Activation Clustering."* In AAAI Artificial Intelligence Safety Workshop, 2018 **(Best Paper)**

Sanjay Purushotham*, **Wilka Carvalho***, Yan Liu. *"Variational Adversarial Deep Domain Adaptation for Health Care Time Series Analysis."* In NeurIPS Machine Learning for Healthcare Workshop, 2016 **(Spotlight)**

Wilka Carvalho. *"Modeling a Detection of internally reflected Cherenkov light (DIRC) Particle*

Detector for High-Multiplicity Collisions." In SURC, 2015

TALKS
Computational Cognitive Science Group (**Invited**). Princeton. (Fall, 2023)
RL at Harvard (**Invited**). Harvard. (August, 2023)
Kempner Institute. Harvard. (March, 2023)
Computational Neuroscience Group. UCL. (October, 2022)
Intelligent Robot Lab (**Invited**). Brown. (February, 2022)
Computational Cognitive Neuroscience Lab. Harvard. (February, 2022)
Computational Cognitive Neuroscience Lab (**Invited**). University of California—Berkeley. (January, 2022)
Deep Learning Community of Practice (**Invited**). Intel. (January, 2022)
Deep RL Workshop (**Contributed Opinion**). NeurIPS. (December, 2021)
Object Representations for Learning Workshop (**Invited**). NeurIPS. (December, 2020)
Cognitive Tools Lab. University of California—San Diego. (November, 2020)
Cognitive Science Community. University of Michigan. (November, 2020)
Stony Brook Society of Physics. Stony Brook, NY. (April, 2020)
Machine Learning Lunch Seminar. University of Southern California. (April, 2017)

PATENTS
Wilka Carvalho. *"Compositional Generalization in Reinforcement Learning."* Under Review

Bryant Chen, **Wilka Carvalho**, Heiko Ludig, Ian Molloy, Jialong Zhang, Benjamin Edwards. *"Detecting poisoning attacks on neural networks by activation clustering."* 2020

Wilka Carvalho, Bryant Chen, Benjamin Edwards, Taesung Lee, Ian Molloy, Jialong Zhang. *"Using Gradients to Detect Backdoors in Neural Networks."* 2018

Wilka Carvalho, Yan Liu, Tanachat Nilanon, Sanjay Purushotham. *"Effective Knowledge Transfer Among Patient Populations via Deep Learning."* 2017

RESEARCH EXPERIENCE
University of Michigan–Ann Arbor, Ann Arbor, Michigan USA
AI Lab, *August 2018 – Present*
Advisors: Honglak Lee, Satinder Singh, Richard Lewis
- Studying how deep reinforcement learning agents can discover representations and temporally-extended behaviors that enable generalization

DeepMind, London, UK
Cognition Team, *September 2022 – May 2023*
Advisor: Daniel Zoran
- Studying transfer in 3D environments with successor features

DeepMind, London, UK
Cognition Team, *March 2021 – July 2021*
Advisor: Murray Shanahan
- Studying agent-state representations for generalization of deep reinforcement learning agents in 3D environments

Microsoft Research, Redmond, Washington USA
Medical Devices Group, *June 2018 – August 2018*
Advisor: Sumit Basu
- Developed a machine learning model that can predict clinical measures from physiological signals measured by a wearable device.

IBM Research, San Jose, California USA

AI Platform Research Group, *September 2017 – December 2017*
Advisor: Heiko Ludwig
- Contributed to novel research algorithm by suggesting subspace projection technique that increased our performance from 15% to 95% accuracy.
- Research led to workshop publication and was presented in global meeting to IBM executives including CEO and selected for funding.

Visa Research, Palo Alto, California USA
Data Analytics Group, *June 2017 – August 2017*
Advisor: Hao Yang
- Implemented model and baselines for language generation and question answering.

University of Southern California, Los Angeles, California USA
Melady Machine Learning Lab, *November 2015 – May 2017*
Advisor: Yan Liu
Samsung and NSF funded project: "Variational Adversarial Deep Domain Adaptation for Health Care Time Series Analysis"
- Implemented novel neural network that employed variational inference and adversarial training for transfer learning of multivariate time-series.
- Research led to a publication and a patent.
- Communicated research to general public through research feature by the USC Graduate School and to technical audience at ICLR poster presentation.

Stony Brook Univeristy, Stony Brook, New York USA
Heavy Ion Research Group, *January 2013 – August 2015*
Advisor: Axel Drees
DOE funded project: "Modeling a Detection of internally reflected Cherenkov light Particle Detector for High-Multiplicity Collisions"
- Contributed methods from multivariate calculus and linear algebra to particle detection algorithm. Accuracy improved from 60% to 80%.
- Designed and implemented a statistical analysis pipeline in C++ for measuring efficacy of particle detection algorithm.

Stony Brook Univeristy, Stony Brook, New York USA
Computational Neuroscience Group, *Fall 2014*
Advisor: Giancarlo La Camera
NSF LSAMP funded project: "Spectral Analysis of Rodent Neural Data"
- Performed spectral analyses on neural data to determine behavioral correlates of neural activity.

California Institute of Technology, Pasadena, California USA
Emotion and Social Cognition Laboratory, *Summer 2014*
Advisor: Ralph Adolphs
HHMI funded project: "Modeling the Cognitive Process of Attributing Traits to Others"
- Formulated a trait learning behavioral experiment to study human inference.

University of Minnesota, Minneapolis, Minnesota USA
Neuromodulation Research and Technology Laboratory, *Summer 2013*
Advisor: Matthew Johnson
NIH funded project: "Modeling Deep Brain Stimulation of Globus Palidus Internus"
- Implemented python script to build a biologically feasible computational model of neural networks

National Central University, Jhongli City, Taiwan
Turbulent Combustion Laboratory, *Summer 2012*
Advisor: Shenqyang Shy

TEACHING EXPERIENCE	**Stony Brook University**, Stony Brook, NY **Calculus Instructor**, *Spring 2015* Worked with two math professors to develop and teach a supplementary calculus curriculum that promoted minority representation in stem majors. **Stony Brook University**, Stony Brook, NY **Educational Opportunity Program Personal Tutor**, *Spring 2013 - Fall 2014* Tutored marginalized students in introductory physics and math courses
SERVICE	Reviewer, Neural Computation, 2023 Reviewer, NeurIPS, 2020, 2021, 2022, 2023 Student Volunteer, ICLR, 2017
OUTREACH	*HS2 Engineer's Highschool Panel*, Virtual, 2021 *Stony Brook Society of Physics*, Stony Brook, NY, 2020 *Michigan Explore Graduate Studies Workshop*, Ann Arbor, MI, 2019 *Research and Fellowships Week NSF Panel*, Los Angeles, CA, 2016 *National Society of Black Engineers Grad Panel*, Los Angeles, CA, 2016 *Graduate School External Fellowship Boot Camp*, Los Angeles, CA, 2016 *Mentored marginalized high school youth through the Pullias Center for Higher Education*, Los Angeles, CA, 2016 *Engineering Graduate Diversity Symposium*, Los Angeles, CA, 2015 *Black Student Association: What it takes to go to Graduate School*, Los Angeles, CA, 2015 *Collegiate Science and Technology Entry Program Undergraduate Research Panel*, Stony Brook, CA, 2014
SKILLS	Machine Learning Software: Pytorch, TensorFlow, Theano, Keras Neuroscience Software: Neuron Languages: Python, C++, C, Java Systems: Unix, Linux, OSX
PRESS	Finding Gaps in Your Grad School Apps This Week In Machine Learning (TWIML) Research Feature University of Michigan Research Feature Exploring the source of social stereotypes Black History Month: Why a career in science? Research Feature by the USC Graduate School 2015 NSF Graduate Research Fellow Wilka Carvalho Biomath Learning Center Launches Modified Supplemental Instruction Program URECA Research of the Month: Wilka Carvalho Student Feature by Stony Brook University
INTERESTS	● traveling ● chess ● software development ● improvisational dance ● deadpan humor