DILIP ARUMUGAM

dilip.a@cs.princeton.edu

PROFESSIONAL POSITIONS

Princeton University
Postdoctoral Research Associate in Computer Science — *July 2024 - present*
Advisor: Thomas L. Griffiths

EDUCATION

Stanford University
Ph.D. in Computer Science — *June 2024*
M.S. in Statistics — *March 2024*
Advisor: Benjamin Van Roy
Thesis: *Deciding What to Learn in Complex Environments*
Committee: Chelsea Finn, Emma Brunskill, Tsachy Weissman, Art Owen (chair)

Brown University
M.S. in Computer Science — *May 2018*
B.S. in Computer Science — *May 2017*
Advisor: Michael L. Littman

TEACHING EXPERIENCE (⋆ = TA, † = HEAD TA, ‡ = GUEST LECTURER)

Stanford University [All Teaching Reviews]

- † Reinforcement Learning (CS234), Emma Brunskill, Spring 2023-2024.
- ⋆ Deep Reinforcement Learning (CS224R), Chelsea Finn & Karol Hausman, Spring 2022-2023.
- ‡ Reinforcement Learning: Frontiers (MS&E338), Benjamin Van Roy, Spring 2022-2023.
- † Reinforcement Learning (CS234), Emma Brunskill, Winter 2022-2023.
- ‡ Reinforcement Learning: Frontiers (MS&E338), Benjamin Van Roy, Spring 2021-2022.
- † Reinforcement Learning (CS234), Emma Brunskill, Winter 2021-2022.
- ⋆ Deep Multi-Task & Meta Learning (CS330), Chelsea Finn & Karol Hausman, Autumn 2021-2022.
- ⋆ Reinforcement Learning (CS234), Emma Brunskill, Winter 2020-2021.
- ⋆ Deep Multi-Task & Meta Learning (CS330), Chelsea Finn, Autumn 2020-2021.
- ⋆ Artificial Intelligence: Principles & Techniques (CS221), Chelsea Finn & Nima Anari, Spring 2019-2020.

Brown University

- ⋆ Deep Learning (CSCI1470/CSCI2470), Eugene Charniak, Fall 2017.
- ⋆ Deep Learning Seminar (CSCI2950K), Eugene Charniak, Fall 2016.
- ⋆ Applied Artifical Intelligence (CSCI1410), Stefanie Tellex, Fall 2015.

PUBLICATIONS

Peer-Reviewed (Journals & Conferences)

· *Bayesian Reinforcement Learning under Limited Cognitive Load*

Dilip Arumugam*, Mark K. Ho*, Noah D. Goodman, Benjamin Van Roy.

Open Mind: Discoveries in Cognitive Science, 2024.

· *Deciding What to Model: Value-Equivalent Sampling for Reinforcement Learning*

Dilip Arumugam, Benjamin Van Roy.

Advances in Neural Information Processing Systems (NeurIPS), 2022.

Multi-disciplinary Conference on Reinforcement Learning & Decision Making (RLDM), 2022.

ICML Workshop on Decision Awareness in Reinforcement Learning, 2022.

· *Planning to the Information Horizon of BAMDPs via Epistemic State Abstraction*

Dilip Arumugam, Satinder Singh.

Advances in Neural Information Processing Systems (NeurIPS), 2022.

NeurIPS Workshop on Ecological Theory of Reinforcement Learning, 2021.

· *The Value of Information When Deciding What to Learn*

Dilip Arumugam, Benjamin Van Roy.

Advances in Neural Information Processing Systems (NeurIPS), 2021.

· *Deciding What to Learn: A Rate-Distortion Approach*

Dilip Arumugam, Benjamin Van Roy.

International Conference on Machine Learning (ICML), 2021.

· *Flexible and Efficient Long-Range Planning through Curious Exploration*

Aidan Curtis, Minjian Xin, **Dilip Arumugam**, Kevin Feigelis, Daniel Yamins

International Conference on Machine Learning (ICML), 2020.

· *Value Preserving State-Action Abstractions*

David Abel, Nathan Umbanhowar, Khimya Khetarpal, **Dilip Arumugam**, Doina Precup, Michael L. Littman.

International Conference on Artificial Intelligence and Statistics (AISTATS), 2020.

Multi-disciplinary Conference on Reinforcement Learning & Decision Making (RLDM), 2019.

ICLR Workshop on Structures and Priors in Reinforcement Learning, 2019.

· *State Abstraction as Compression in Apprenticeship Learning*

David Abel, **Dilip Arumugam**, Kavosh Asadi, Yuu Jinnai, Michael L. Littman, Lawson L.S. Wong.

Association for the Advancement of Artificial Intelligence (AAAI) Conference, 2019.

· *Grounding Natural Language Instructions to Semantic Goal Representations for Abstraction and Generalization*

Dilip Arumugam*, Siddharth Karamcheti*, Nakul Gopalan, Eddie Williams, Mina Rhee, Lawson L.S. Wong, Stefanie Tellex.

Autonomous Robots (AuRo), 2018.

· *State Abstractions for Lifelong Reinforcement Learning*

David Abel, **Dilip Arumugam**, Lucas Lehnert, Michael L. Littman.

International Conference on Machine Learning (ICML), 2018.

· *Sequence-to-Sequence Language Grounding of Non-Markovian Task Specifications*

Nakul Gopalan*, **Dilip Arumugam***, Lawson L.S. Wong, Stefanie Tellex.

Robotics: Science and Systems, 2018.

· *Accurately and Efficiently Interpreting Human-Robot Instructions of Varying Granularities*

Dilip Arumugam*, Siddharth Karamcheti*, Nakul Gopalan, Lawson L.S. Wong, Stefanie Tellex.

Robotics: Science and Systems, 2017.

· *Grounding English Commands to Reward Functions*

James MacGlashan, Monica Babes-Vroman, M. desJardins, Michael L. Littman, Smaranda Muresan, Shawn Squire, Stefanie Tellex, **Dilip Arumugam**, and Lei Yang.

Robotics: Science and Systems, 2015.

Manuscripts & Preprints

· Stay Tuned!

Non-Archival Publications

· *Exploration Unbound*

Dilip Arumugam*, Wanqiao Xu*, Benjamin Van Roy.

RLC Finding the Frame Workshop, 2024.

· *Satisficing Exploration for Deep Reinforcement Learning*

Dilip Arumugam, Saurabh Kumar, Ramki Gummadi, Benjamin Van Roy.

RLC Finding the Frame Workshop (Oral), 2024.

· *Social Contract AI: Aligning AI Assistants with Implicit Group Norms*

Jan-Philipp Fränken, Sam Kwok, Peixuan Ye, Kanishk Gandhi, **Dilip Arumugam**, Jared Moore, Alex Tamkin, Tobias Gerstenberg, Noah D. Goodman.

NeurIPS Workshop on Socially Responsible Language Modelling Research (Spotlight + Oral), 2023.

· *Cultural Reinforcement Learning: A Framework for Modeling Cumulative Culture on a Limited Channel*

Ben Prystawski, **Dilip Arumugam**, Noah D. Goodman.

Proceedings of the 45th Annual Meeting of the Cognitive Science Society (CogSci), 2023.

· *Hindsight-DICE: Stable Credit Assignment for Deep Reinforcement Learning*

Akash Velu*, Skanda Vaidyanath*, **Dilip Arumugam**.

Preprint, 2023.

· *Fine-Tuning Inclusive Language Models*

Wanqiao Xu, Shi Dong, **Dilip Arumugam**, Benjamin Van Roy.

Preprint, 2023.

· *Inclusive Artificial Intelligence*

Dilip Arumugam, Shi Dong, Benjamin Van Roy.

Preprint, 2022.

· *On Rate-Distortion Theory in Capacity-Limited Cognition & Reinforcement Learning*

Dilip Arumugam, Mark K. Ho, Noah D. Goodman, Benjamin Van Roy.

NeurIPS Workshop on Information-Theoretic Principles in Cognitive Systems, 2022.

· *In the ZONE: Measuring Difficulty and Progression in Curriculum Generation*

Rose E. Wang, Jesse Mu, **Dilip Arumugam**, Natasha Jaques, Noah D. Goodman.

NeurIPS Deep Reinforcement Learning Workshop, 2022.

· *Bad-Policy Density: A Measure of Reinforcement Learning Hardness*

David Abel, Cameron Allen, **Dilip Arumugam**, D. Ellis Hershkowitz, Michael L. Littman, Lawson L.S. Wong.

ICML Workshop on Reinforcement Learning Theory, 2021.

· *An Information-Theoretic Perspective on Credit Assignment in Reinforcement Learning*

Dilip Arumugam, Peter Henderson, Pierre-Luc Bacon.

NeurIPS Workshop on Biological and Artificial Reinforcement Learning, 2020.

· *Randomized Value Functions via Posterior State-Abstraction Sampling*

Dilip Arumugam, Benjamin Van Roy.

NeurIPS Workshop on Biological and Artificial Reinforcement Learning, 2020.

· *Reparameterized Variational Divergence Minimization for Stable Imitation*

Dilip Arumugam, Debadeepta Dey, Alekh Agarwal, Asli Celikyilmaz, Elnaz Nouri, Bill Dolan.

Preprint, 2019.

· *Goal-Directed Learning as a Bi-level Optimization Problem*

Pierre-Luc Bacon, **Dilip Arumugam**, Emma Brunskill

Multi-disciplinary Conference on Reinforcement Learning & Decision Making (RLDM), 2019.

· *Deep Reinforcement Learning from Policy-Dependent Human Feedback*

Dilip Arumugam, Jun Ki Lee, Sophie Saskin, Michael L. Littman.

Preprint, 2018.

· *A Tale of Two DRAGGNs: A Hybrid Approach for Interpreting Action-Oriented and Goal-Oriented Instructions*

Siddharth Karamcheti, Eddie Williams, **Dilip Arumugam**, Mina Rhee, Nakul Gopalan, Lawson L.S. Wong, Stefanie Tellex

ACL Workshop on Language Grounding for Robotics, 2017. [*Best Paper Award*]

· *Mitigating Planner Overfitting in Model-Based Reinforcement Learning*

Dilip Arumugam, David Abel, Kavosh Asadi, Nakul Gopalan, Chris Grimm, Jun Ki Lee, Lucas Lehnert, Michael L. Littman.

Preprint, 2017.

INDUSTRY EXPERIENCE

Technical Advisor on AI — Summer 2023
SCB10X, Bangkok, Thailand
Contact: Mukaya (Tai) Panich

Research Scientist Intern — Summer 2021
DeepMind, London, UK
Mentors: Brendan O'Donoghue & Satinder Singh

Visiting Researcher — Summer 2020
Montreal Institute for Learning Algorithms (MILA), Montreal, QC
Mentor: Pierre-Luc Bacon

Reinforcement Learning Group Research Intern — Summer 2019
Microsoft Research, Redmond, WA
Mentor: Debadeepta Dey

Reinforcement Learning Group Research Intern — Summer 2016
Microsoft Research, Cambridge, UK
Mentors: Matthew Johnson, Katja Hofmann, & Dave Bignell

INVITED TALKS

Deciding What to Learn in Complex Environments — 2024

- Brown University
- Harvard University
- University of Toronto
- Princeton University
- Stanford University

Deciding What to Learn in Multi-Armed Bandits — 2023-2024

- Arizona State University
- Princeton University
- University of Southern California
- Cornell University
- Cornell Tech
- University of Washington
- University of Chicago

AWARDS & SERVICE

Conference Reviewer

ICML – 2020 [Top Reviewer Award], 2021, 2022, 2023, 2024

NeurIPS – 2021 [Outstanding Reviewer Award], 2023 [Top Reviewer Award]

ICLR – 2022

RLC – 2024

CoRL – 2018

IEEE Transactions on Information Theory – 2024

Journal Reviewer

JMLR – 2023

JAIR – 2023

IEEE TPAMI – 2021

Program Committee

NeurIPS Deep RL Workshop – 2020, 2021, 2022

ICML Decision Awareness in RL Workshop – 2022

ICML Neural Compression Workshop – 2023

NeurIPS Information-Theoretic Principles in Cognitive Systems Workshop – 2023

Student Co-organizer, Stanford Reinforcement Learning Group, 2020 - 2021.
Brown University CS Department Senior Prize, 2017.
Sigma Xi Scientific Research Honor Society, 2017.
Brown University CS Department Undergraduate Research Symposium Award, Spring 2017.