dasgupta.ishita@gmail.com
ishita-dg.github.io

Ishita Dasgupta

Researcher in machine learning and computational cognitive science.

PROFESSIONAL EXPERIENCE

DeepMind, New York City

Senior Research Scientist — Nov 2022 – Present

Research Scientist — Dec 2020 – Oct 2022

- Using foundation models in embodied environments with high-dimensional observation and action spaces.
  - Co-led a research team of 6 as part of a 50+ person long-term research effort.
  - Collaborating with language models for embodied reasoning. Dasgupta *et al.* NeurIPS Language and Reinforcement Learning (LaReL) 2022; **best paper award**.
  - Distilling internet-scale vision-language models into embodied agents. Sumers, ..., Dasgupta. ICML 2023.
- Reasoning and deliberation in large language models.
  - Data generation and evaluations for multimodal reasoning in mainline Gemini models.
  - Language models show human-like content effects on reasoning. Dasgupta*, Lampinen* et al. arXiv 2022.
  - Can language models learn from explanations in context? Lampinen, Dasgupta et al. EMNLP 2022.
- Understanding transformer models.
  - Transformers generalize differently from information stored in context vs in weights. Dasgupta*, Chan* et al., NeurIPS Memory in Artificial and Real Intelligence (MemARI) 2022.
  - Are CNNs or Transformers more like human vision? Tuli, Dasgupta et al. CogSci 2021.
- Further details in academic CV and on Google Scholar.
- First Research Scientist hire at DeepMind NYC; played a central role in building the team.

Princeton University. Postdoctoral Fellow, Dept. of Computer Science. Jan – Dec 2020

– Analyzing and augmenting representations learned by AI systems, focus on inductive bias & abstraction.

– 8 publications (3 first author), including a NeurIPS Outstanding Main Track Paper award.

EDUCATION

Harvard University, Ph.D. in Physics. March 2020

Thesis: *Algorithms for ecological rationality in humans and machines.*

Indian Institute of Technology Bombay, B.Tech. with Honours in Engineering Physics. August 2014

RECENT AWARDS

NeurIPS Outstanding Main Track Paper. 2022

Using natural language and program abstractions to instill human inductive biases in machines.

Best Paper Award, NeurIPS Language and Reinforcement Learning Workshop. 2022

Collaborating with language models for embodied reasoning.

SKILLS

Machine learning methods; NLP, CV; large-scale training and serving.
Online crowd-sourcing platforms: experiment design, collection, and analysis of human behavioral data.
Technical leadership (tech lead for a research team of 5+), mentorship (supervised 5+ junior researchers).

SERVICE

Area Chair (ICLR 2024, NeurIPS 2023, NeurIPS workshops 2022, 2021), extensive peer review.
Volunteering at Harvard Women in Physics, Teach for India Mumbai, DeepMind Scholars.