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

# Ishita Dasgupta

Researcher in machine learning and computational cognitive science.

## PROFESSIONAL EXPERIENCE

**DeepMind, New York City**

Research Scientist (Senior Research Scientist since Nov 2022) Dec 2020 – Present

- Technical leadership and internal impact.
  - Co-led a research team of 6 within a 50+ person long-term research effort on adapting large foundation models for use in environments with high-dimensional action and observations spaces.
  - Defined novel visual reasoning capabilities for Gemini foundation models; owned data generation and evaluation for these capabilities.
  - Played a central role in hiring and building the DeepMind NYC team (first Research Scientist hire).
- Research impact: Representative publications by topic below; more on Google Scholar.

  Using foundation models in embodied environments.
  - 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.
  - 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 representations in transformers.
  - 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.

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

- Analyzed and augmented 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

Computational models and behavioral experiments for human decision making. 12 publications, 3 awards.

**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; extensive experiments with large-scale models.
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.