

Karthik Viswanathan

✉ vkarthik095@gmail.com • 🏠 Website • 🎓 Scholar • 🔗 LinkedIn • 🐙 GitHub

Education

University of Amsterdam

PhD in Physics

Thesis: *From Language Models to Cosmic Structures: A Geometric Perspective*

Advisors: Dr. Matteo Biagetti and Dr. Jan Pieter van der Schaar

Amsterdam

2021 – 2025

University of Amsterdam

MSc in Physics and Astronomy (Theoretical Physics), cum laude

Thesis: *Exploring the Spectral Theory/Topological Strings duality*

Advisor: Dr. Marcel Vonk

GPA: 8.7/10

Amsterdam

2019 – 2021

Indian Institute of Technology Madras

B.Tech in Engineering Physics

Thesis: *Real Space Renormalization and Applications to Machine Learning*

Advisor: Dr. Ashwin Joy

GPA: 8.29/10

Chennai

2013 – 2017

Publications

TopoFisher: Learning Topological Summary Statistics by Maximizing Fisher Information

Matteo Biagetti, Mathieu Carrière, Francesco Conti, Enrico Maria Ferrari, Sven Heydenreich, Karthik Viswanathan

Under review [[paper](#), [code](#)]

A Transformer Architecture Alteration to Incentivise Externalised Reasoning

Elizabeth Pavlova*, Mariia Koroliuk*, Karthik Viswanathan*, Cameron Tice, Edward Young, Puria Radmard

Preprint [[paper](#), [code](#)]

Probing Geometry of Next Token Prediction Using Cumulant Expansion of Softmax Entropy

Karthik Viswanathan*, Sang Eon Park*

ICML 2025: Workshop on High-dimensional Learning Dynamics [[paper](#), [code](#)]

Persistent Topological Features in Large Language Models

Yuri Gardinazzi*, Karthik Viswanathan*, Giada Panerai, Alessio Ansuini, Alberto Cazzaniga, Matteo Biagetti

ICML 2025 (Poster) [[paper](#), [code](#)]

The Intrinsic Dimension of Prompts in Internal Representations of Large Language Models

Karthik Viswanathan*, Yuri Gardinazzi*, Giada Panerai, Alberto Cazzaniga, Matteo Biagetti

Transactions on Machine Learning Research (TMLR), 2026 [[paper](#), [code](#)]

Cosmology with persistent homology: A Fisher forecast

Jacky Yip, Matteo Biagetti, Alex Cole, Karthik Viswanathan, Gary Shiu

Journal of Cosmology and Astroparticle Physics (JCAP), Vol. 2024, Issue 09 [[paper](#), [code](#)]

Academic Activities

Verified Mechanisms

London

Research lead, independent research project

Aug 2026 – present

LLM agents propose and formally verify theorems about internal computations in transformers.

Secured project funding in the grantmaking.ai Launch Round [[blogpost](#), [code](#), [grant](#)]

ML Alignment & Theory Scholars Program (MATS 9)

Berkeley & London

Mentors: Adam Shai & Paul Riechers (Simplex AI Safety)

Jan 2026 – present

Do Alignment Pretraining if you want your Linear Probes to Work Better [[poster](#)]

Mentorship for Alignment Research Students (MARS 3.0)

Cambridge

Mentors: Puria Radmard, Cameron Tice & Edward Young (Geodesic Research)

Jul – Sep 2025

Architectures for Increased Externalisation of Reasoning [[blogpost](#), [code](#)]

Workshop: Interpretability in LLMs using Geometric and Statistical methods

Amsterdam

Organized by Karthik Viswanathan, Jan Pieter van der Schaar

May 2025

Reviewer:

- Physical Review Letters
- Principled Design for Trustworthy AI (ICLR 2026 Workshop)
- Mechanistic Interpretability Workshop (ICML 2026 Workshop)

Professional Experience

Goldman Sachs

Bangalore

Surveillance Analyst

2017 – 2019

Developed ML models for anomaly detection in financial data

Manager: Prof. Howard Karloff

Goldman Sachs

Bangalore

Summer Intern

2016

Implemented fast Personalized PageRank in a MapReduce framework

Mentor: Dr. Koushik Balasubramanian

Awards & Distinctions

Sander Bais Prize for Academic Merit

Amsterdam

Institute for Theoretical Physics Amsterdam

2020

For exceptional academic performance in the master's program

Honorable Mention

Marrakech

ACM ICPC World Finals

2015

Represented India in the international collegiate programming contest.

Participant

Bangalore

International Olympiad in Informatics Training Camp

2013

Selected after ranking in top 25 in the Indian National Olympiad in Informatics

Organization & Mentoring

LASR Labs Extension

Mentor

Jun 2026 – present

Project: Detecting CoT Unfaithfulness with Attention Probes

Berkeley Research Academy

Research Mentor

2026

Project: How Many Attention Heads Compute a Boolean Function? A Formally Verified Analysis

AREA Science Park, Trieste

Master Student Supervisor

2024

Student: Giada Panerai

Thesis: Zig-Zag Persistence in Neural Networks Representations

University of Amsterdam

Master Student Mentor

2023

Student: Sibilla Bouche

Thesis: The persistence of non-Gaussian features: a Neural Ratio Estimation approach

University of Amsterdam

Master Student Supervisor

2023

Student: Sliem el Ela

Thesis: From Primordials to Persistence: A Dual Exploration of Multiparameter Topology and Cosmic Origins

University of Amsterdam

Journal Club Organizer

2022

Cosmology Journal Club

Research Visits

2024: AREA Science Park, Trieste, Italy

September 16 – October 4

2024: AREA Science Park, Trieste, Italy

February 15 – April 27

Technical Skills

Programming: PyTorch, TensorFlow, Hadoop, MapReduce, Mathematica

HPC Resources: LUMI consortium, Snellius supercomputer

Teaching

University of Amsterdam

Teaching Assistant

MSc Physics Program

2020 – 2024

Courses: Topological Data Analysis: a Physics Perspective, Advanced Cosmology, ML for Physics and Astronomy, QFT in Curved Spacetimes, General Relativity

References

Dr. Matteo Biagetti

AREA Science Park, Trieste

matteo.biagetti@areasciencepark.it

Dr. Jan Pieter van der Schaar

University of Amsterdam

j.p.vanderschaar@uva.nl

Dr. Alberto Cazzaniga

AREA Science Park, Trieste

alberto.cazzaniga@areasciencepark.it

Hannes Whittingham

Meridian Cambridge

hannes@meridiancambridge.org