Arvindh Arun

🌐 arvindh75.github.io   @ iiith@arvindharun.com   ⚙ github.com/arvindh75   🎓 Google Scholar
in linkedin.com/in/arvindh75   🐦 twitter.com/arvindh__a

Education

Current **Aug 2024**	**ELLIS and IMPRS-IS PhD (University of Stuttgart & University of Edinburgh)** **Stuttgart, Germany** PhD in Artificial Intelligence › Advised by Prof. Steffen Staab and Prof. Antonio Vergari › Working on Foundation Models for Knowledge Graphs, LLM Evaluations, Graph-based Retrieval Augmented Generation for LLMs
May 2024 **Aug 2019**	**International Institute of Information Technology (IIIT H)** **Hyderabad, India** B.Tech. + MS by Research, Computer Science & Engineering › Cumulative GPA - 9.07/10.00 › Dean's Academic List - 2020-21 M (Top 10%), 2021-22 M (Top 15%), 2022-23 S (Top 5%) › Dean's Research List - 2022-23 (Top 5%) › Represented the College Football Team, 5x medals in Track & Field, 3x medals in Inter-house football

Experience

Aug 2024 **Jan 2024**	**Google DeepMind | Information Networks (InfoNet) Team** **Remote** *Student Research Collaborator* › Worked on leveraging topological information from user interactions to gather insight into the quality of user-generated content › Extended Vietoris Rips Complex and Persistent Images to post profiling on social networks › Introduced a new benchmark to evaluate controversy detection models mimicking the real-world data distribution
Jul 2023 **May 2023**	**JPMorgan Chase | Corporate Investment Banking** **Bangalore, India** *Applied AI/ML Intern* › Worked on mining interesting insights from transactional databases for Root Cause Analysis. › Utilized streaming frameworks such as PySpark to optimize the workflow. › Proposed a new framework utilizing unsupervised GNNs inspired by the CTR prediction domain. › Received a Pre-placement Full-time offer for my contributions.
Aug 2022 **Jun 2022**	**The Chinese University of Hong Kong | Department of Systems Engineering** **Hong Kong** *Research Intern | Advisor: Prof. Hong Cheng* › Worked on the domain of Clustering/Community detection in Dynamic social networks and studying their properties. › Focused on minimizing the time complexity and the space complexity of the existing real-time solutions. › Explored various state-of-the-art solutions and built a visualization tool to study their efficacy.
May 2024 **May 2021**	**Precog Research Group** [🌐] **Hyderabad, India** *Undergraduate Researcher | Advisor: Prof. Ponnurangam Kumaraguru "PK"* › Analyzing social networks and other large networks primarily using their network information. › **Georgia Institute of Technology, Atlanta** Collaboration with **Prof. Polo Chau** on analyzing the Quantitative and Qualitative impact of fraudulent behavior on social networks and their identification. › **Adobe Research, Bangalore** Collaboration with **Dr. Ramasuri Narayanam** on exploring the landscape of group fairness in unsupervised contrastive learning based GNNs. › **Indiana University, Bloomington** Collaboration with **Prof. Jisun An** on analyzing the changes to the hate speech dynamics after the liberalization of social media platforms.
May 2021 **Jan 2021**	**NemoCare** **Bangalore, India** *Cross Platform Development Intern* › Built the first prototype of the Android and iOS app for this Shark Tank funded, Google accelerated, Asia Hardware Battle 2019-winning startup led by a Forbes 30 under 30 recipient. › Integrated the app with their pre-existing AWS backend ecosystem and their hardware by leveraging Lambda and EC2.

Publications

[8] SEMMA: A Semantic Aware Knowledge Graph Foundation Model [PDF]
Arvindh Arun, Sumit Kumar, Mojtaba Nayyeri, Bo Xiong, Ponnurangam Kumaraguru, Antonio Vergari and Steffen Staab
The 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP)   **[EMNLP '25 (Main)]**

[7] A Cognac shot to forget bad memories: Corrective Unlearning in GNNs [PDF]
Varshita Kolipaka, Akshit Sinha, Debangan Mishra, Sumit Kumar, Arvindh Arun*, Shashwat Goel*, and Ponnurangam Kumaraguru (* = Equally Advising)
42nd International Conference on Machine Learning (ICML)   **[ICML '25]**

[6] `Best Paper` **Topo Goes Political: TDA-Based Controversy Detection in Imbalanced Reddit Political Data** [PDF]
Arvindh Arun* Karuna K Chandra*, Akshit Sinha, Balakumar Velayutham, Jashn Arora, Manish Jain, and Ponnurangam Kumaraguru (* = Equal Contribution)
5th International Workshop on Computational Methods for Online Discourse Analysis @ WWW   **[BeyondFacts @ WWW '25]**

[5] Sanity Checks for Evaluating Graph Unlearning [PDF]
Varshita Kolipaka, Akshit Sinha, Debangan Mishra, Sumit Kumar, Arvindh Arun, Shashwat Goel, and Ponnurangam Kumaraguru
3rd Conference on Lifelong Learning Agents Workshop Track (CoLLAS)   **[CoLLAS-W '24]**

[4] X-posing Free Speech: Examining the Impact of Moderation Relaxation on Online Social Networks [PDF]
Arvindh Arun*, Saurav Chhatani*, Jisun An, and Ponnurangam Kumaraguru (* = Equal Contribution)
8th Workshop on Online Abuse and Harms (WOAH) at NAACL   **[WOAH @ NAACL '24]**

[3] `Oral` **CAFIN: Centrality Aware Fairness inducing IN-processing for Unsupervised Representation Learning on Graphs** [PDF | code]
Arvindh Arun, Aakash Aanegola, Amul Agrawal, Ramasuri Narayanam, and Ponnurangam Kumaraguru
26th European Conference on Artificial Intelligence (ECAI)   **[ECAI '23]**

[2] Erasing Labor with Labor: Dark Patterns and Lockstep Behaviors on the Google Play Store [PDF]
Ashwin Singh, Arvindh Arun, Pulak Malhotra, Pooja Desur, Ayushi Jain, Duen Horng Chau, and Ponnurangam Kumaraguru
LBR Track at 33rd ACM Conference on Hypertext and Social Media (ACM HT)
`Media` *Work covered by Montreal AI Ethics Institute*   **[LBR @ HT '22]**

[1] Ensembling Pre-Trained Language Models with Feature Engineering [PDF | code]
T.H. Arjun, Arvindh Arun, and Ponnurangam Kumaraguru
GermEval 2021 Shared Task on the Identification of Toxic, Engaging, and Fact-Claiming Comments   **[GermEval '21]**

Honours, Awards and Test Scores

Y Combinator AI Startup School, 2025 [🌐]   Accepted to attend a funded 2-day AI startup school in the Bay Area.

Google Research Week, 2024 [🌐]   Accepted to attend a funded 3-day ML research event, hosted by Google Research India.

Microsoft Research Travel Grant, 2023 [🌐]   Awarded $2500 for attending ECAI'23 held in Krakow, Poland.

Stellantis NV Hackathon, 2022   Finished as a runner-up in both rounds, winning a cash prize worth 125,000 INR.

Facebook Hackercup, 2022   Reached Round 2, finished Top 10% globally.

GermEval, 2021   Ranked 6th globally on cumulative task rank.

PicoCTF by CMU, 2021   Finished 102nd, Top 2% globally.

HowzHack (DreamVu), 2019   Finished 2nd for the problem statement exploring the applications of computer vision for Football's Video Assistant Referees (VAR).

UGEE, 2019 20th rank in India.

NextGenius Scholarship, 2019   Finalist (withdrew) selected for fully funded Undergraduate programme at University Of The South, Tennessee, USA.

Teaching

Foundations of Machine Learning (UoS)   *Teaching Assistant*   April '25 - June '25
> Introductory ML course for undergrads.
> Covered lectures on Linear Algebra.

Deep Learning Lab (UoS)   *Course Co-instructor*   Oct '24 - Feb '25
> Introductory lab course for Deep Learning with a special focus on Knowledge Graphs, Diffusion models, and LLMs.
> Covered lectures on Graph Learning and Good practices in Deep Learning.

Computational Social Science (IIITH - CS4.301)   *Teaching Assistant*   Jan '23 - May '23
> Course covers social computing concepts and social media analysis.
> Responsibilities included curating assignments, conducting weekly tutorials, and grading quizzes and assignment submissions.

Data & Applications (IIITH - CS9.435)   *Teaching Assistant*   Aug '22 - Dec '22
> Course covers the basics of database systems and introduces SQL.
> Responsibilities included curating assignments, conducting weekly tutorials, and grading quizzes and assignment submissions.

Academic Service

Reviewer   ACL ARR October '24, ICLR '25

Selected Research Projects

Ensemble of Task-Specific Language Models for Brain Encoding [Arxiv | ⊙]   Jan '23 - May '23
Cognitive Science and AI (CS9.432)

We improve the performance of brain encoders by creating an ensemble model out of 10 popular Language Models (2 syntactic and 8 semantic). Going by the hypothesis that different LMs capture different forms of semantics better, we beat the current baselines by 10% on average across all ROIs through our ensembling methods.

Privacy Policy QnA bot [⊙]   Aug '22 - Dec '22
Online Privacy (CS4.407)

We fine-tuned the Jurassic-1 language model on a new dataset consisting of top Indian companies' Privacy Policies for QA and Summarization tasks. Also deployed a Telegram bot for easy access by connecting it with our hosted model.

Wikipedia Open QnA [PDF | ⊙]   Aug '21 - Dec '21
Advanced NLP (CS7.501)

We implemented the DrQA Retriever and Reader, and extracted features required to run the RankQA model. We also indexed a large Wikipedia data dump to train on and tested it on the popular SQuAD dataset.