

Arvindh Arun

[arvindh75.github.io](https://github.com/arvindh75) [@iiith@arvindharun.com](mailto:iiith@arvindharun.com) github.com/arvindh75 [Google Scholar](https://scholar.google.com/citations?user=...)
[in linkedin.com/in/arvindh75](https://www.linkedin.com/in/arvindh75) twitter.com/arvindh__a

Education

May 2024 Aug 2019	International Institute of Information Technology (IIIT H) B.Tech. + MS by Research, Computer Science & Engineering <ul style="list-style-type: none">> 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	Hyderabad, India
----------------------	--	------------------

Experience

Present Jan 2024	DeepMind (formerly Google Research) InfoNet <i>Student Research Collaborator</i> Working on leveraging topological information from user interactions to gather insight on the quality of user-generated content.	Remote
Jul 2023 May 2023	JPMorgan Chase Corporate Investment Banking <i>Applied AI/ML Intern</i> <ul style="list-style-type: none">> 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.	Bangalore, India
Aug 2022 Jun 2022	The Chinese University of Hong Kong Department of Systems Engineering <i>Research Intern Advisor: Prof. Hong Cheng</i> <ul style="list-style-type: none">> 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.	Hong Kong
May 2024 May 2021	Precog Research Group  <i>Undergraduate Researcher Advisor: Prof. Ponnurangam Kumaraguru "PK"</i> <ul style="list-style-type: none">> 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.	Hyderabad, India
May 2021 Jan 2021	NemoCare <i>Cross Platform Development Intern</i> Built the first prototype of the Android and iOS app for this Google accelerated, Asia Hardware Battle 2019-winning startup led by a Forbes 30 under 30 recipient.	Bangalore, India

Publications

- [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, Mexico City, Mexico [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, Krakow, Poland [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, Barcelona, Spain
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

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

TOEFL – 115/120

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, 2020 Ranked 6th globally on cumulative task rank.

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

Google Hashcode, 2020 Finished 457th.

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.

JEE, 2019 Top 1.9% in India.

CBSE Class 12, 2019 Top 1.5% in India.

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

Teaching

Computational Social Science (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 (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.

Selected Research Projects

Ensemble of Task-Specific Language Models for Brain Encoding [PDF | 🌐] 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.