

# Darpan Jain

GitHub: [github.com/darpan-jain](https://github.com/darpan-jain)  
LinkedIn: [linkedin.com/darpan-jain](https://www.linkedin.com/in/darpan-jain)

Email: [darpanjainn@gmail.com](mailto:darpanjainn@gmail.com)

Website: <https://darpanjain.com>

Mobile: +1(213)-522-6633

## EDUCATION

**Master of Science, Electrical and Computer Engineering**, University of Southern California May 2023  
- Specialization in Machine Learning and Data Science — GPA - 3.92

**Bachelor of Engineering, Electrical Engineering**, University of Pune Apr 2018

## WORK EXPERIENCE

Pixstory, **Lead Machine Learning Engineer** May 2023-Present

- Leading Pixstory's Retrieval Augmented Generation (RAG)-based Conversational Search, serving accurate and nuanced user query responses with verified knowledge base grounding.
- Mitigating hallucination and misinformation in generation by Large Language Models (LLMs) by rigorous vetting and re-ranking trusted sources for reliable information retrieval.

Information Sciences Institute, **Applied Researcher** Jan 2022-July 2023

- Led the development of conflict resolution dialogue agents for DARPA's Civil Sanctuary program as an Applied Researcher at the Natural Language group ([CUTE LAB NAME](#)) at USC-ISI.
- Designed and deployed a scalable dialogue agent for mitigating toxic behavior on social media using non-violent communication strategies, including controlled text generation utilizing GPT-3.5's advanced transformer-based language generation capabilities and prompting techniques. Read the [research paper here](#).
- Moderated over 15k posts with an 85% true positive rate in French and German on popular subreddits.

Lumin.ai, **Software Development Intern** May 2022-August 2022

- Contributed to developing the conversational AI for Lumin.ai as a Smart Scheduler and Sales Accelerator to enhance customer interactions and optimize the sales funnel for 5+ product and franchise owners.
- Advocated for efficient documentation using Markdown and managed customer interactions for multiple clients.

Warner Bros Discovery

- Software Development Engineer - Machine Learning** Sep 2020-Aug 2021
  - Boosted user engagement on Discovery Plus (D+) as part of the "Personalization and Recommendations" team.
  - Employed Apache Spark and Airflow to build efficient ETL pipelines in a distributed environment and executed A/B tests to validate and refine new recommendation features for the D+ platform.
  - Led the development of server-side ad-insertion SDKs for Android and Web, driving increased ad revenue for D+ sports content.
- Machine Learning Engineer**, (formerly AdSparx Inc.) Mar 2019-Sep 2020
  - Led the research, design, and development of MiDAS, a micro-service-based system serving personalized ads to millions of users on OTT platforms.
  - Designed and trained Computer Vision models with TensorFlow, deployed on Kubernetes to enable dynamic server-side ad insertion for 40+ channels, boosting revenue for US publishers.
  - Member of the senior development team at AdSparx which Warner Bros Discovery acquired in September 2020.

## PUBLICATIONS

[Can Language Model Moderators Improve the Health of Online Discourse?](#)

Hyundong Cho, Shuai Liu, Taiwei Shi, [Darpan Jain](#), Basem Rizk, Yuyang Huang, Zixun Lu, Nuan Wen, Jonathan Gratch, Emilio Ferrara, Jonathan May

## CONSULTING EXPERIENCE

**Machine Learning Lead**, Defeat Covid-19 Mar 2020-May 2020

- Co-founded a non-profit aimed at providing support to authorities and the public during the COVID-19 pandemic, with features such as real-time monitoring dashboards, identification of high-risk areas, and a multilingual QA chatbot.

**Artificial Intelligence Consultant**, Prasaurus Sports Analytics Jun 2019-Aug 2019

- Spearheaded the development of a Badminton analytics system to provide in-depth game analysis, using activity recognition and player tracking to produce player heat-maps and dominant areas metrics.

**Computer Vision Engineer**, Occipital Tech Feb 2019-Mar 2019

- Developed real-time image pre-processing solutions using Background Segmentation for enhancing blurry and under-lit images, achieving 14-15% noise reduction and 25% faster processing time for high-quality results.

## TECHNICAL SKILLS

**Languages:** Python, GoLang, Javascript, Java, Scala, C++, SQL, R

**NLP & Machine Learning:** PyTorch, Tensorflow, Scikit-learn, Langchain, LlamaIndex, NLTK, SpaCy, CoreNLP, Hugging Face Transformers, Sagemaker, MLflow, Kubeflow

**Frameworks & Technologies:** Git, Docker, Kubernetes, Terraform, Jenkins, REST, Linux, AWS, GCP, Azure

**Databases & Monitoring tools:** Vector Stores, SQL, Kafka, Spark, Elasticsearch, DynamoDB, Redis, Splunk, Datadog