

# Gaurav Arora

MACHINE LEARNING ENGINEER @ HAPTIK, EX - GOLDMAN SACHS

medium.com/@giganticgemmic | contactgauravforwork@gmail.com | goru001.github.io | goru001 | gaurav-arora-23593220

## Experience

### Haptik (part of Jio Platforms Limited)

Mumbai

MACHINE LEARNING ENGINEER - 2

July 2019 - Present

- Working with the ML team on fundamental Conversational-AI problems using **Deep Learning**
- Built the **Intent Detection** System for Haptik's NLU Engine, owning it from Research to Production
- The performance of new system was bench-marked against 10 data-sets, including 3 open source datasets and it was **25% more accurate** than their previous system, with latency under **150ms**
- We also benchmarked the new system against popular NLU platforms like Dialogflow, RASA, LUIS in the paper **HINT3: Raising the bar for Intent Detection in the Wild** and Haptik's system achieved similar performance as others.
- The new system is underpinned by key concepts like **Transfer Learning**, **Data-Augmentation**, protection against **Adversarial Queries** and is like an **Auto-ML** service for bot developers

### Goldman Sachs

Bengaluru

ANALYST

June 2018 - July 2019

- Worked with the UX and Productivity team on **Analytics** for Desktop Assistant, which is a productivity tool used by every employee in the firm to perform activities like Call, IM, Email etc
- My work at GS brought to the fore key metrics about Product Health, User Behavior which were missing earlier

### Researchshala

Chandigarh

Co-FOUNDER

Nov. 2016 - Mar. 2018

- Researchshala is an online platform to connect professors to research-interns, helping them with their research projects
- Researchshala had more than 40 research projects and 90+ students working under professors from universities like Carnegie Mellon University, IITs etc. during that time

### Goldman Sachs

Bengaluru

INTERN

Jan. 2017 - June 2017

- Worked with the team that built and maintains the firm's standard UI Development Framework - **the UI Toolkit**

## Selected Publications

### Natural Language Toolkit for Indic Languages - iNLTK

goru001/inltk

ACCEPTED AT EMNLP-2020'S NLP-OSS WORKSHOP

<https://arxiv.org/abs/2009.12534>

- iNLTK aims to provide out of the box support for various NLP tasks that an application developer might need for Indic languages
- iNLTK provides out-of-the-box support for **Data Augmentation**, **Textual Similarity**, **Sentence Embeddings**, **Word Embeddings**, **Tokenization** and **Text Generation** in 13 Indic Languages
- iNLTK has **40,000+** downloads, **600+** stars on GitHub, and has been widely shared and appreciated on Twitter, LinkedIn, Reddit
- By using pre-trained models and data augmentation from iNLTK, we achieve more than 95% of the previous best performance by using less than 10% of the training data on publicly available classification datasets

### HINT3: Raising the bar for Intent Detection in the Wild

hellohaptik/HINT3

ACCEPTED AT EMNLP-2020'S INSIGHTS WORKSHOP

<https://arxiv.org/abs/2009.13833>

- This paper introduces 3 new datasets created from real user queries received by chatbots, evaluates popular NLU platforms and highlights critical gaps in language understanding
- Paper introduces novel **subset approach** for evaluation and discusses trade-off between out-of-domain vs in-scope performance

### Pre-training ULMFiT on Synthetically Generated Code-Mixed Data for Hate Speech Detection

goru001/nlp-for-manglish,nlp-for-tanglish

ACCEPTED AT DRAVIDIAN CODEMIX HASOC-2020 @ FIRE-2020

<https://arxiv.org/abs/2010.02094>

- This paper proposes pre-training ULMFiT on synthetically generated code-mixed data, generated by modelling code-mixed data generation as a Markov process
- Model achieved 0.88 weighted F1-score for code-mixed Tamil-English language in Sub-task B and got 2nd rank on the leader-board

## Honors & Awards

---

### CODING

Mar. 2020	<b>Fast.ai International Fellow</b> , for contributions to Fast.ai forums	Worldwide
Mar. 2019	<b>Fast.ai International Fellow</b> , for contributions to Fast.ai forums	Worldwide
	<b>Top-17% , Human Protein Atlas Image Classification, Kaggle</b> , for developing Deep Learning	
Dec.2018	model which classified mixed patterns of proteins in microscope images. The competition had 2172 teams, but I participated individually and hence had 100% contribution in the 366 <sup>th</sup> placed solution	Worldwide
Oct.2017	<b>1st Prize,IEEE-Hackathon</b> , for developing chat-bot to help people with emotional decisions in life	PEC
Feb.2016	<b>Top-100</b> , among 500,000 students in IT-Olympiad,2016.	India
Oct.2016	<b>2nd-Prize, IEEE-Hackathon</b> , for developing an Augmented reality application to help teachers	PEC
Mar.2016	<b>All India Rank-6</b> , in IEEE Programming League, among over 1200 undergraduate students.	India
Mar.2016	<b>2nd Rank</b> , CodeWars,a competitive-programming event hosted by IEEE,PEC on CodeChef	PEC

### ACADEMIC

Nov.2016	<b>Research Scholarship</b> , for Personal Emotional Doctor - Bot	PEC
2014	<b>All India Rank-885</b> , JEE-Mains, among 1.4 million candidates	India
2014	<b>1st Rank-Opener,PEC</b> , for <b>best</b> JEE-Mains rank among 600 students of the session 2014-2018.	PEC
2014	<b>1 Lakh Scholarship</b> , from CBSE for 96.4% marks in 12th Boards and 10 CGPA in 10th.	India
2014	<b>Letter of Appreciation</b> , from HRD Ministry,Govt. of India for 96.4% in CBSE-12th exams	India
2011	<b>Catch Them Young</b> , Was among the <b>top-40</b> students selected from tricity by INFOSYS for 2-week Programming-Basics training on their campus	INFOSYS

## Education

---

### International Institute of Information Technology (IIIT - Hyderabad)

Bengaluru, India

ADVANCED CERTIFICATION IN ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

2018 - 2019

- This was in-depth and comprehensive AI-ML program for tech professionals. I was certified by the program with "Very good" performance.

### Punjab Engineering College

Chandigarh, India

BACHELOR OF TECHNOLOGY (B.TECH.), COMPUTER SCIENCE AND ENGINEERING

2014-2018

- Best JEE-Mains Rank in class
- CGPA : 7.77/10

### GMSSS-16

Chandigarh, India

INTERMEDIATE

2014

- 12th standard : 96.4%
- 10th standard(CGPA) : 10/10

## Skills & Courses

---

#### Mathematics

Discrete Structures for Computer Science, Vector Calculus, Fourier Series and Laplace Transform, Operations Research, Mathematics for Machine Learning: Linear Algebra and Multivariate Calculus (Coursera)

#### Computer Science

Data Structures and Algorithms, Computer Architecture and Organization, OOP, Microprocessor, DBMS, Operating Systems, Computer Networks, Theory of Computation, Artificial Intelligence, Computer Graphics, Mobile Computing, Fastai: Part 1 and Part 2, DeepLearning.ai by Andrew Ng, Deep Learning by Prof. Mitesh Khapra, IIT Madras

#### Programming & Web Machine Learning tools

Python, C, C++, Javascript, TypeScript, EcmaScript6, AngularJS, ReactJS, Angular4, Webpack, Django with Python Pytorch, Pandas, Numpy, ScikitLearn, SciPy, Fastai, Transformers library

## Extracurricular Activity

---

MOTIVATIONAL-SPEAKER

2014 - PRESENT

- Given more than 10 motivational lectures at schools, college etc to an audience of about 100, talking about *student-life challenges* and *success-hacks*.